MATHEMATICAL METHODS IN SCIENCE

GEORGE PÓLYA

The great book of Nature lies ever open before our eyes and the true philosophy is written in it... But we cannot read it unless we have first learned the language and the characters in which it is written.... It is written in mathematical language and the characters are triangles, circles, and other geometrical figures...

Galileo: Saggiatore, Opere VI, p. 232

The Mathematical Association of America
New Mathematical Library

MATHEMATICAL METHODS IN SCIENCE

by

George Pólya

Stanford University

Edited by

Leon Bowden

University of Victoria, B.C.

26

THE MATHEMATICAL ASSOCIATION OF AMERICA

This booklet originally appeared in 1963 as Volume XI of the Studies in Mathematics series sponsored by the School Mathematics Studies Group which was financially supported by the National Science Foundation.

The Mathematical Association of America expresses its gratitude to SMSG, and specially to its Director, E. G. Begle, for granting permission to have a revised edition of this valuable work published as Volume 26 of MAA's New Mathematical Library series.

Drawings by Buehler & McFadden

Second Printing

Published in Washington by the Mathematical Association of America

Library of Congress Catalog Card Number: 76-25863

Complete Set ISBN 0-88385-600-X

Vol. 26 0-88385-626-3

Manufactured in the United States of America

Note to the Reader

This book is one of a series written by professional mathematicians in order to make some important mathematical ideas interesting and understandable to a large audience of high school students and laymen. Most of the volumes in the *New Mathematical Library* cover topics not usually included in the high school curriculum; they vary in difficulty, and, even within a single book, some parts require a greater degree of concentration than others. Thus, while the reader needs little technical knowledge to understand most of these books, he will have to make an intellectual effort.

If the reader has so far encountered mathematics only in classroom work, he should keep in mind that a book on mathematics cannot be read quickly. Nor must he expect to understand all parts of the book on first reading. He should feel free to skip complicated parts and return to them later; often an argument will be clarified by a subsequent remark. On the other hand, sections containing thoroughly familiar material may be read very quickly.

The best way to learn mathematics is to *do* mathematics. The reader is urged to acquire the habit of reading with paper and pencil in hand; in this way mathematics will become increasingly meaningful to him.

The authors and editorial committee are interested in reactions to the books in this series and hope that readers will write to: Anneli Lax, Editor, New Mathematical Library, NEW YORK UNIVERSITY, THE COURANT INSTITUTE OF MATHEMATICAL SCIENCES, 251 Mercer Street, New York, N. Y. 10012.

The Editors

NEW MATHEMATICAL LIBRARY

1 Numbers: Rational and Irrational *by Ivan Niven*
2 What is Calculus About? *by W. W. Sawyer*
3 An Introduction to Inequalities *by E. F. Beckenbach and R. Bellman*
4 Geometric Inequalities *by N. D. Kazarinoff*
5 The Contest Problem Book I Annual High School Mathematics Examinations 1950–1960. Compiled and with solutions *by Charles T. Salkind*
6 The Lore of Large Numbers *by P. J. Davis*
7 Uses of Infinity *by Leo Zippin*
8 Geometric Transformations I *by I. M. Yaglom, translated by A. Shields*
9 Continued Fractions *by Carl D. Olds*
10 Graphs and Their Uses *by Oystein Ore*
11 Hungarian Problem Books I and II, Based on the Eötvös
12 Competitions 1894–1905 and 1906–1928, *translated by E. Rapaport*
13 Episodes from the Early History of Mathematics *by A. Aaboe*
14 Groups and Their Graphs *by I. Grossman and W. Magnus*
15 The Mathematics of Choice *by Ivan Niven*
16 From Pythagoras to Einstein *by K. O. Friedrichs*
17 The Contest Problem Book II Annual High School Mathematics Examinations 1961–1965. Compiled and with solutions *by Charles T. Salkind*
18 First Concepts of Topology *by W. G. Chinn and N. E. Steenrod*
19 Geometry Revisited *by H. S. M. Coxeter and S. L. Greitzer*
20 Invitation to Number Theory *by Oystein Ore*
21 Geometric Transformations II *by I. M. Yaglom, translated by A. Shields*
22 Elementary Cryptanalysis—A Mathematical Approach *by A. Sinkov*
23 Ingenuity in Mathematics *by Ross Honsberger*
24 Geometric Transformations III *by I. M. Yaglom, translated by A. Shenitzer*
25 The Contest Problem Book III Annual High School Mathematics Examinations 1966–1972. Compiled and with solutions *by C. T. Salkind and J. M. Earl*
26 Mathematical Methods in Science *by George Pólya*
27 International Mathematical Olympiads—1959–1977. Compiled and with solutions *by S. L. Greitzer*
28 The Mathematics of Games and Gambling *by Edward W. Packel*
29 The Contest Problem Book IV Annual High School Mathematics Examinations 1973–1982. Compiled and with solutions *by R. A. Artino, A. M. Gaglione and N. Shell*
30 The Role of Mathematics in Science *by M. M. Schiffer and L. Bowden*

Other titles in preparation

Preface

"Mathematical Methods in Science" is the title of a course which I have given several times at Stanford University to teachers, or prospective teachers, of mathematics and science. The following pages present those chapters of the course the contents of which were not incorporated in a previously printed work. (See *Mathematics and Plausible Reasoning*, Vol. 1, especially chapters III, VIII, and IX.)

The following presentation is due to Professor Leon Bowden of the University of Victoria, who carefully followed in substance a tape recording of the course, but added several details and several picturesque sentences of his own. Some peculiarities of the oral presentation have been preserved: a certain broadness and some traces of improvisation.

One of the essential tendencies of the course is to point to the history of certain elementary parts of science as a source of efficient teaching in the classroom. Several historical details are somewhat distorted: some intentionally, to bring them down to the level of the high school, but a few details may be unintentionally distorted, I am afraid. A careful confrontation of the pedagogically appropriate with the historically correct version would be most desirable, but was not feasible within the limits of time and energy at my disposal. A few non-historical niceties are also somewhat roughly treated, for reasons of space and pedagogy.

I hope that the following pages will be useful, yet they should not be regarded as a finished expression of the views offered.

My warmest thanks go to Professor Bowden.

George Pólya

Stanford University, July, 1963

Preface to the Revised Edition

The present edition leaves the original essentially unchanged. However, several misprints and minor slips have been corrected, details of the presentation improved, parts of subsections 1.2.6, 3.1.3, 3.4.3, 5.1.1, and 5.1.3 have been rewritten, and a final section "What is a Differential Equation?" has been added.

In the light of classroom experience (my own experience, that of Professor Bowden and several colleagues) it was advisable not to deviate from the original general conception of the work. The preface to the original edition mentioned the difficulty of reconciling the pedagogically appropriate with the historically correct. Also this somewhat delicate balance was left undisturbed.

Professor Bowden and I are deeply grateful to Professor Anneli Lax for incorporating this revised version into the New Mathematical Library, for numerous improvements, both major and minor, of the text, and for editorial expertise which greatly facilitated the toil and trouble of revision.

George Pólya

Stanford University, October 1976

Contents

Introduction

In these lectures we will discuss:

(1) Very simple physical or pre-physical problems; problems that could be discussed at the high school level.

(2) The relation of mathematics to science and of science to mathematics. This relation is a two-way street. Though more usual, it is not always the case that mathematics is applied to science; also there is traffic in the opposite direction. Good driving takes note of the oncoming traffic.

(3) Elementary calculus, for without some calculus one's idea of how mathematics is applied to science is necessarily inadequate.

Also, as their title indicates, these lectures will deal with my ideas about methods. First, let me say that there is no one teaching method which is *the* method; there are as many good methods as there are good teachers. To teach effectively a teacher must develop a feeling for his subject; he cannot make his students sense its vitality if he does not sense it himself. He cannot share his enthusiasm when he has no enthusiasm to share. How he makes his point may be as important as the point he makes; he must personally feel it to be important; he must develop his personality.

In my presentation I shall, by and large, follow the genetic method. The essential idea of this method is that the order in which knowledge has been acquired by the human race will be a good order for its acquisition by the individual. The sciences came in a certain order; an order determined by human interest and inherent difficulty. Mathematics and astronomy were the first sciences really worthy of the name; later came mechanics, optics, and so on. At each stage of its development the human race has had a certain climate of opinion, a way of looking, conceptually, at the world. The next glimmer of fresh understanding had to grow out of what was already understood. The next move forward, halting shuffle, faltering step, or stride with some confidence, was developed upon how well the race could then walk. As for the human race, so for the human child. But this is not to say that to teach science we must repeat the thousand and one errors of the past, each ill-directed shuffle. It is to say that the sequence in which the major strides forward were made is a good sequence in which to teach them. The genetic method is a guide to, not a substitute for, judgment.

CHAPTER ONE

From the History of Astronomy: Measurement and Successive Approximation

SECTION 1. MEASUREMENT

1.1.1 The Tunnel

Astronomers have measured the distance of the Sun from the Earth; even the distance of the fixed stars. How did they do this? Not by strolling through outer space with a measuring rod. The distance of places that cannot be reached is calculated from the distance of places that can be reached. To measure the stars we get down to Earth; cosmological survey has a terrestial base.

We begin with a terrestial problem. Due to increasing population a certain city of ancient Greece found its water supply insufficient, so that water had to be channeled in from a source in the nearby mountains. And since, unfortunately, a large hill intervened, there was no alternative to tunneling. Working from both sides of the hill, the tunnelers met in the middle as planned. See Figure 1.1.

How did the planners determine the correct direction to ensure that the two crews would meet? How would you have planned the job? Remember that the Greeks could not use radio signal or telescope, for they had neither. Nevertheless they devised a method and actually succeeded in making their tunnels from both sides meet somewhere inside the hill. Think about it.

Of course, had not the source been on a higher level than the city, there would not have been gravity to make the water flow through this aqueduct. But, to better concentrate on the crux of the matter, let us

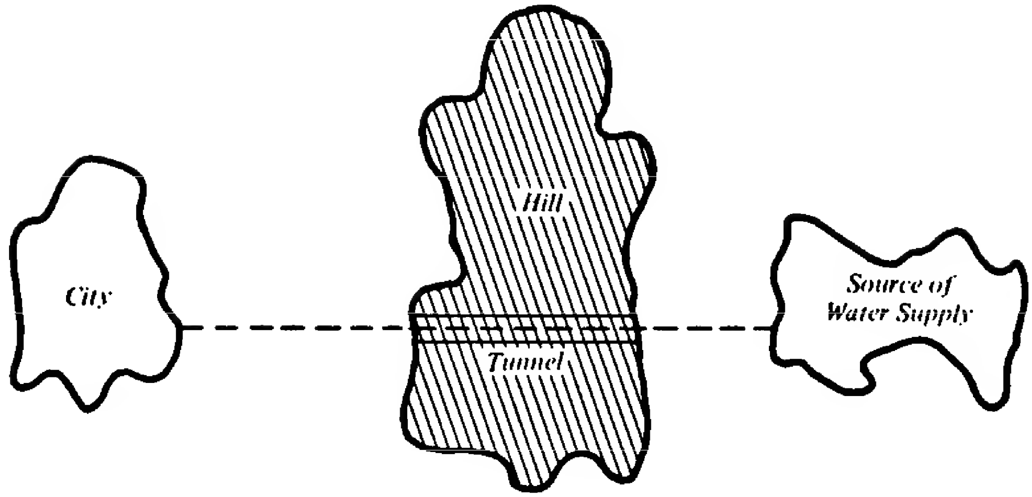

Figure 1.1

neglect the complication due to difference of levels. Essentially the problem is this. How do we determine the would-be line of sight between two coplanar points C and S when a hill intervenes? See Fig. 1.2. Here we have a problem of applied geometry. How are we to construct the segments CC', SS' of the straight line CS without joining C to S? It is not permitted to traverse the shaded area.

That which cannot be connected directly can only be connected indirectly. Let O (O is for *Outside*) be a point from which both C and S are observable. Joining O to C and to S, we have the situation of Fig. 1.3.

Surely this diagram must suggest application of the geometry of the triangle. And how do we specify a triangle? By measuring its angles and

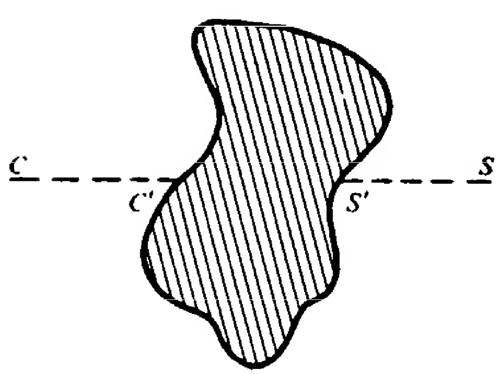

Figure 1.2

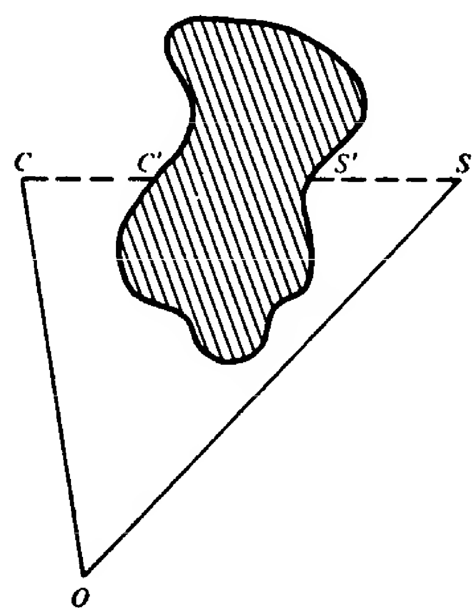

Figure 1.3

sides. And what angles are measurable in Fig. 1.3? The angle at O can be measured, for C and S are both visible from O. But what about the angles at C and S? We cannot measure $\angle OCC'$ since the hill intervenes between C and S, and therefore the direction of CC' is unknown. For the same reason we cannot measure $\angle OSS'$, or the length of CS. Thus the measurables are OC, OS and the angle at O—two sides and included angle—sufficient to specify $\triangle OCS$ uniquely.

Suppose that OC is found to be 2 miles, OS 3 miles, and $\angle COS = 53°$. We can draw a scale model with, say, O_1C_1 20 inches, O_1S_1 30 inches, and, of course, with the included $\angle C_1O_1S_1 = 53°$. And since similar triangles are equiangular, it follows that $\angle OCC'$ (i.e., $\angle OCS$) $= \angle O_1C_1S_1$, and $\angle OSS'$ (i.e., $\angle OSC$) $= \angle O_1S_1C_1$. See Fig. 1.4. The problem is solved.

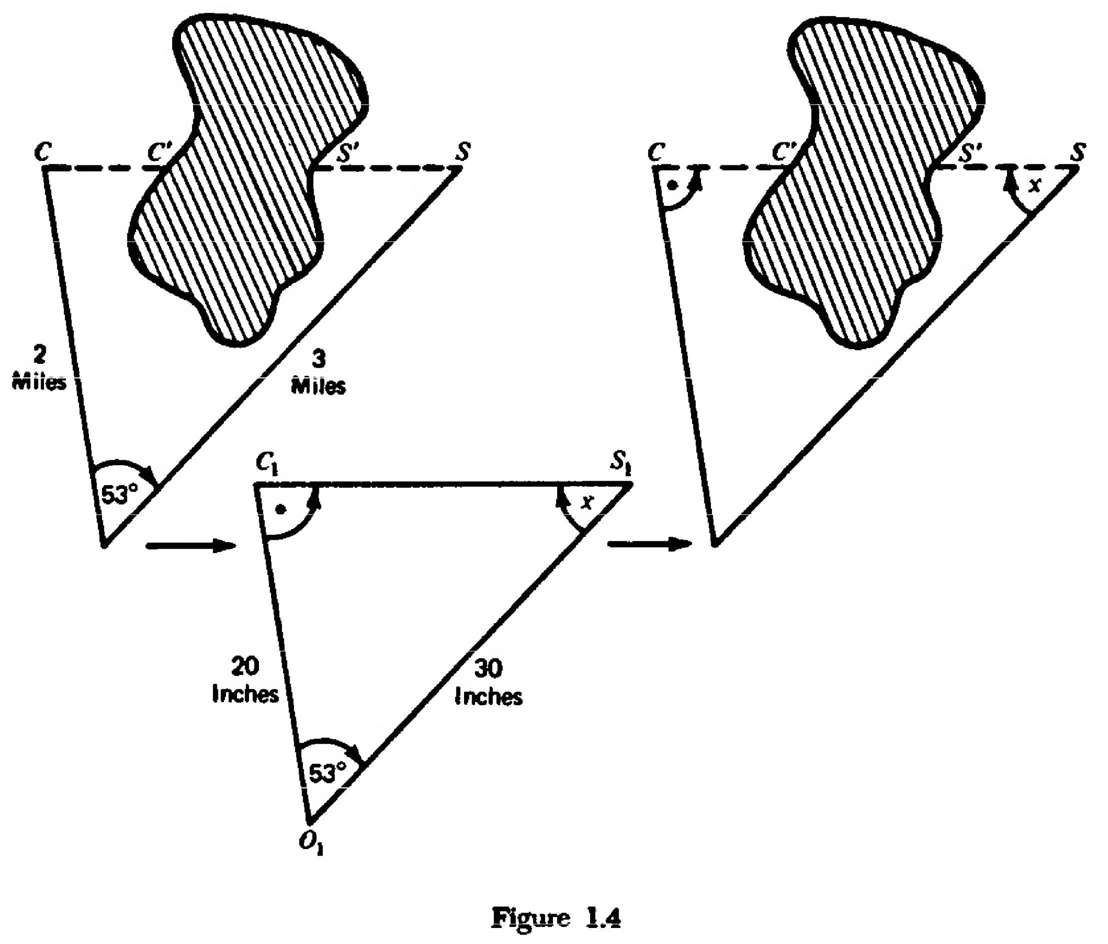

Figure 1.4

The alert reader will have already appreciated that the length of the tunnel, and consequently the amount of tunneling for each crew, is easily deduced. The directions of CC' and SS' having been determined, their lengths can be measured; from the length of C_1S_1 in the auxiliary triangle the length of CS can be deduced by simple proportion: the

length of the tunnel is the difference between the latter and the sum of CC' and SS'.*

1.1.2 Measuring: Triangulating

Next a word about the important practical business of making measurements. How do we measure an angle? We do it basically the same way today as the Greeks did it two thousand years ago. The modern theodolite effects greater precision, it is better built; the principle is no better, it is the same, essentially a protractor. What is a protractor? An arc or the whole circumference of a circle divided into equal parts. See Fig. 1.5. In changing our line of sight from OC to OS, it is rotated through a certain number of subdivisions of the circular arc. Since the amount of turning is proportional to this number, the number is a measure of $\angle COS$. It is conventional from Babylonian times to consider a complete revolution to be 360 degrees, and therefore to divide the whole circumference into 360 equal parts. When greater accuracy is required and the protractor is large enough to allow further division, each part is subdivided into 60 parts to read off sixtieths of a degree (minutes), and, in turn, each such sixtieth is subdivided into 60 parts to read off sixtieths of a minute (seconds).

To measure $\angle COS$ with great accuracy, the lines of sight OC, OS must be precisely known. Precision is achieved by sighting the objects C and S with the aid of a cross hair at the end of a cylindrical tube mounted at O. A modern refinement is the magnification attained by making the tube telescopic. See Fig. 1.6. Yet no matter how refined the refinements, error is inevitable. So today's surveyor just as the surveyor of two thousand years ago, makes several measurements of an angle and takes their average. The measurement of an angle remains a fundamental operation.

The reader who, in trying his hand at amateur carpentry, attempts to make a picture frame without the aid of a miter box knows to his sorrow how difficult it is to make the fourth corner fit. His sad experience may tempt him to suppose that accurate measurement of lengths is easier than that of angles. No, when it comes to surveying, the measurement of an angle is a relatively precise operation. To establish a base line a mile or two long is a difficult (and expensive) operation. It has to be made completely flat. A further difficulty is that measuring rods or chains change length with temperature. Another difficulty is that the line must be straight. The men who built the two-mile-long linear accelerator at Stanford could tell you that measuring angles is much easier than constructing a straight line.

---

*See B. L. van der Waerden; *Science Awakening*; p. 102–104, The tunnel of Samos, also Plate 14, for more, and more correct, historical details.

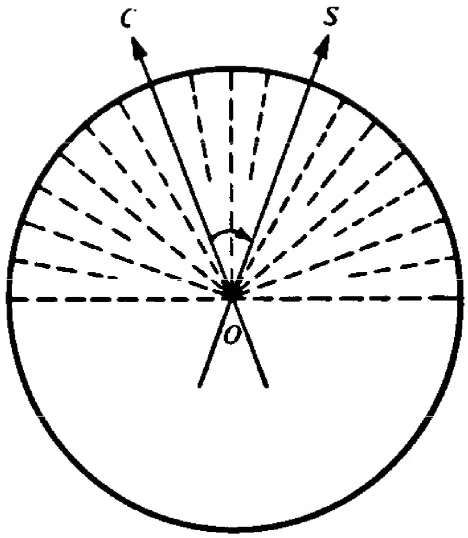

Figure 1.5

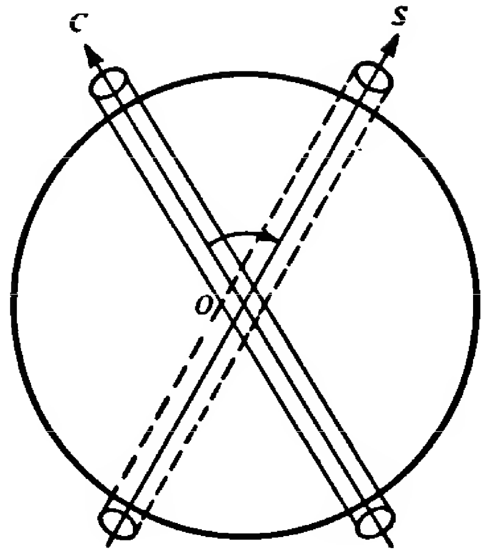

Figure 1.6

When a base line AB is established, the sighting of some prominent distant object C, such as a church steeple or mountain peak, enables angles ABC, BAC to be measured and hence AC, BC computed by trigonometry. These in turn can be used as base lines from which to sight other prominent topographical points C_1, C_2, leading to the use of AC_1, CC_1, CC_2, BC_2 as further base lines, and so on. See Fig. 1.7. In this way, that is, by what is called *triangulation*, a whole country or continent can be surveyed.

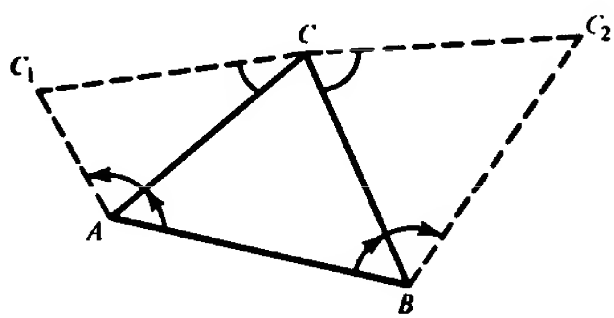

Figure 1.7

1.1.3 How Far Away is the Moon?

From the Earth we turn to the heavens. How are we to measure the distance of the Moon from the Earth? Since this distance cannot be measured directly, it must be measured indirectly; it can only be determined by calculation from accessible distances. So we need a known base line. Basically we have a problem of triangulation. Can the problem be related to that of $\triangle ABC$ of Fig. 1.7? Consider Fig. 1.8. Yes, if we can

determine the straight line distance AB and angles α' and β'. Granted that the Earth is a sphere, if the distance AB on the Earth's surface (the arc length) has been measured and θ is known, then OA can be calculated (or, conversely, if the radius OA is known, then θ can be calculated). Hence by consideration of the isosceles $\triangle OAB$, the straight line distance AB is computed. But how is α' to be determined? $\angle OAB$ can be computed from $\triangle OAB$, so that α' will be known when α is known. But what is α? α is the angle which the line of sight to the Moon makes with the vertical at A. And how is the vertical determined? Yes, by suspending the plumb line. Similarly β' is determined by first measuring β. The problem is indeed related! Note that a base line is indispensable, so that before the Greeks could measure the distance of the Moon from the Earth they had to know the shape and the size (i.e., radius or circumference) of the Earth.*

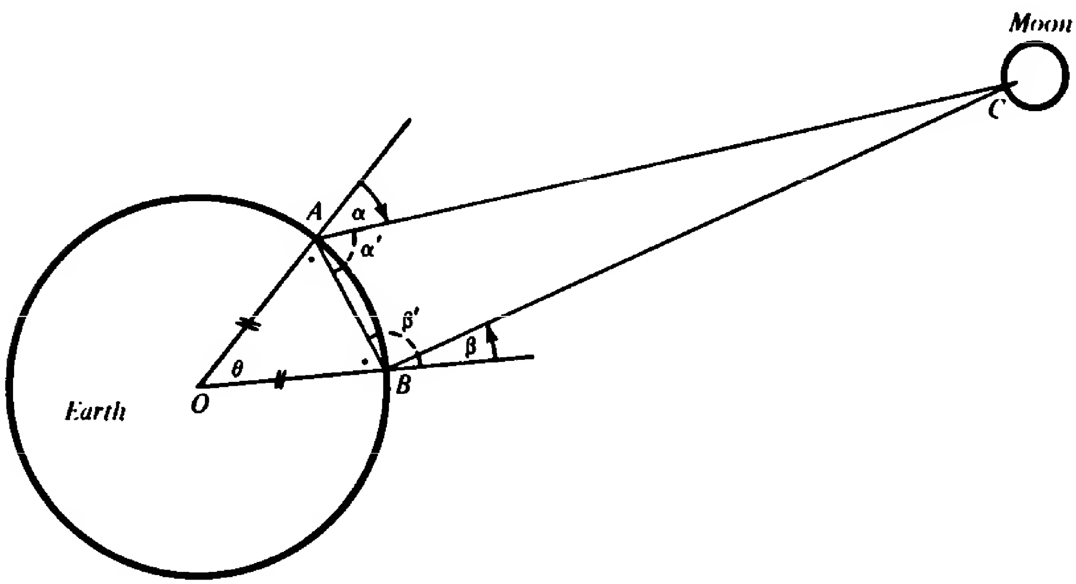

Figure 1.8

One obstacle remains; the Moon moves relatively to the Earth. If β is measured at B after α was measured at A, then β is not the angle to the vertical at B made by the Moon when at C; it is the angle made by the Moon from some subsequent position—say C'. Instead of a triangle with vertices A, B, C, we are confronted with a quadrilateral with vertices A, B, C, C', and the method has failed. For triangulation C, C' must be coincident; α and β must be measured simultaneously.

*If A and B are on the same meridian, and the moon is in the plane of the meridian, the points O, A, B and C are coplanar.

But how is the measurer at B to know when the measurer at A is measuring? To signal to a second measurer just a few miles away a lantern would serve; yet for accurate triangulation such a short base line would not. Remember that AC, BC are each some tens of thousands of miles. Ideally a base line should be of the same order of magnitude; at least it must be hundreds. Remember also that the Greeks had no radio with which to transmit signals, nor had they accurate watches (just clepsydras). Doesn't their problem seem insuperable? Yes; yet they surmounted it. How? Let us for the moment indulge in wishful thinking of a particularly whimsical kind: what a pity the Greeks couldn't get the Man in the Moon to cooperate by signaling! His signal would have been visible at A and B simultaneously. Put less fancifully, measurers had to wait for some happening on the Moon visible from Earth. What happening? A lunar eclipse. See Fig. 1.9. The eclipse provides four distinct events, observable simultaneously from A and B: (1) the beginning and (2) the completion of the Moon's entry of the Earth's shadow, (3) the beginning and (4) the completion of the Moon's emergence from the Earth's shadow. Had you appreciated how useful eclipses are? Compare the idea here with that of O in Fig. 1.3. Isn't human ingenuity a fascinating thing?

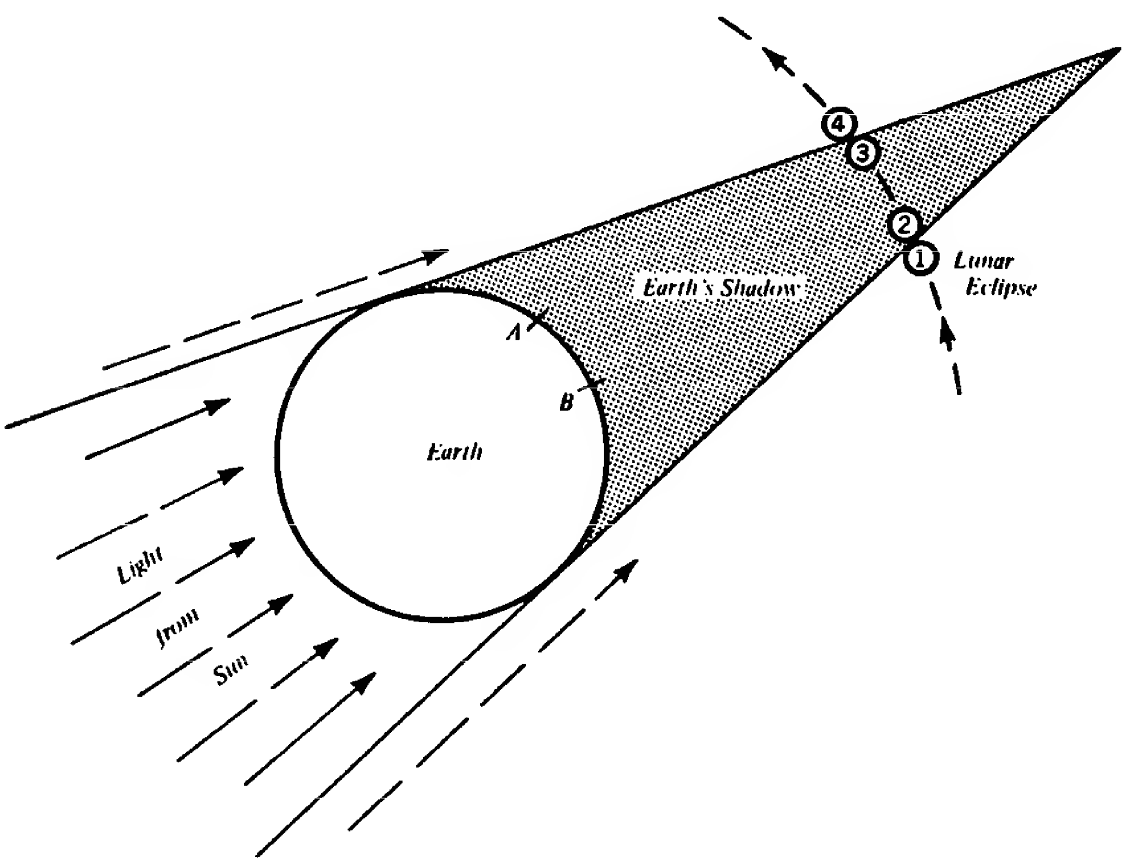

Figure 1.9

1.1.4 Why Teach Triangulation?

Let us for a moment turn from triangulating to teaching. Why should the typical student be interested in those wretched triangles? Hasn't he already genuine interests? Baseball, television, and the girl next door? After all he is only human. Yet precisely because he is human he has human interests—and human curiosity. Why should not the subject be introduced in the way that must interest him? Until he has developed to some level of sophistication he cannot share sophisticated interests. He is to be brought to see that without knowledge of triangles there is no trigonometry; that without trigonometry we put back the clock millennia to Standard Darkness Time and antedate the Greeks.

SECTION 2. ASTRONOMICAL MEASUREMENTS

1.2.1 Aristarchus of Samos

Aristarchus, a famous Greek mathematician and astronomer, was born on the island of Samos about 310 B.C. and died about 230 B.C., so that he was a contemporary of Euclid. His fame rests on his heliocentric theory, the theory that the Earth and planets revolve in orbits around the Sun. Perhaps "theory" is too strong a word, for his proofs were weak; yet it was a great idea, an idea redeveloped centuries later by Copernicus.

Although Aristarchus did not know the distances of the Moon and Sun from the Earth, he was able to estimate their ratio. His method depends upon a most ingenious idea. To better appreciate his ingenuity, stop and ponder awhile. What method would you use? His idea is germinated in an understanding of how the phases of the Moon occur.

Why do we sometimes see a full moon, at other times a half-moon, and when there is a new moon, nothing at all? Because the Moon has no light of its own but depends upon the Sun for its illumination, only one half of its spherical surface is lit up; the other hemisphere is unilluminated. (More precisely, granted the natural assumption that the Sun is a very great distance from the Moon, the beam of its light which illuminates the Moon will be practically a parallel beam and so light up very little more than one hemisphere.) See Fig. 1.10. An observer at P_1 (ideally transparent so as not to block any of the Moon's sunlight) would see almost a whole illuminated hemisphere, i.e., full moon. At P_2 what does he see? His field of vision now includes less of the illuminated hemisphere and a little of the unilluminated—and therefore invisible—hemisphere. He sees the Moon as a shape somewhat larger than a semi-circle. At P_3 his field of vision includes but little of the illuminated part and much of the un-

illuminated. Since only the illuminated is visible he sees the Moon as a crescent. At P_4 his field of vision includes none of the illuminated part, he sees no moon at all—the beginning of the new moon. In what position (relative to the Sun and Moon) would he see precisely a half-moon?

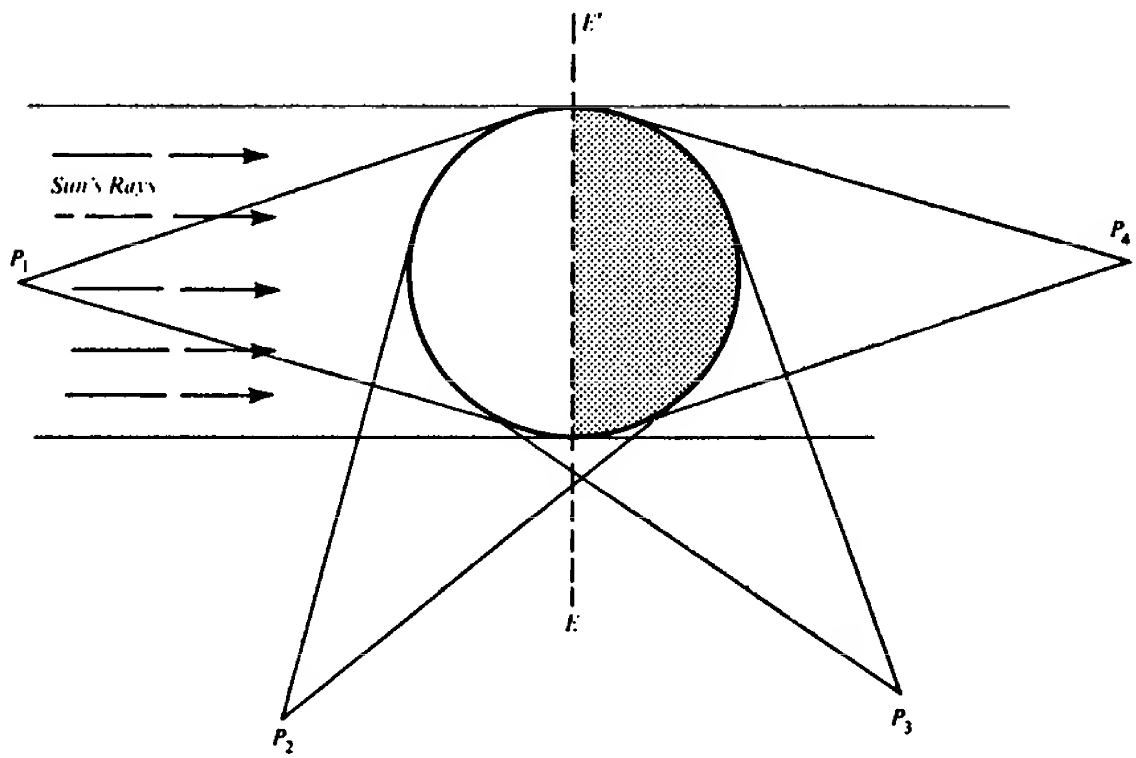

Figure 1.10

Is it not visibly obvious that an observer will have one half of the illuminated and one half of the unilluminated hemispheres in his field of vision, and consequently will see a half-moon, only when he is somewhere on the line EE'? In short, referring to Fig. 1.11, an observer on Earth sees a half-moon only when $\angle EMS$ is a right angle.

Under good atmospheric conditions the Moon is sometimes visible in the daytime, especially near sunset and sunrise. So, sometimes both Sun and Moon are visible. So, sometimes (though less often) both Sun and Moon are visible when the phase of the Moon is half-moon. So? Measure $\angle MES$ on such an occasion, of course. This is what Aristarchus did.

First note that without any measuring at all, since the hypotenuse of a right-angled triangle is the greatest side, we may infer, as did the Greeks, that the Sun is farther from the Earth than the Moon. Next note that when α is measured, the third angle (the complement of α) is determined, so that the shape but not the size of $\triangle EMS$ is known. Consequently, although the actual length of any side is not determinate, the ratio of any pair is. It immediately follows from the definition of

cosine that the ratio of the distances ME (Moon–Earth), SE (Sun–Earth) is given by

$$\frac{ME}{SE} = \cos\alpha.$$

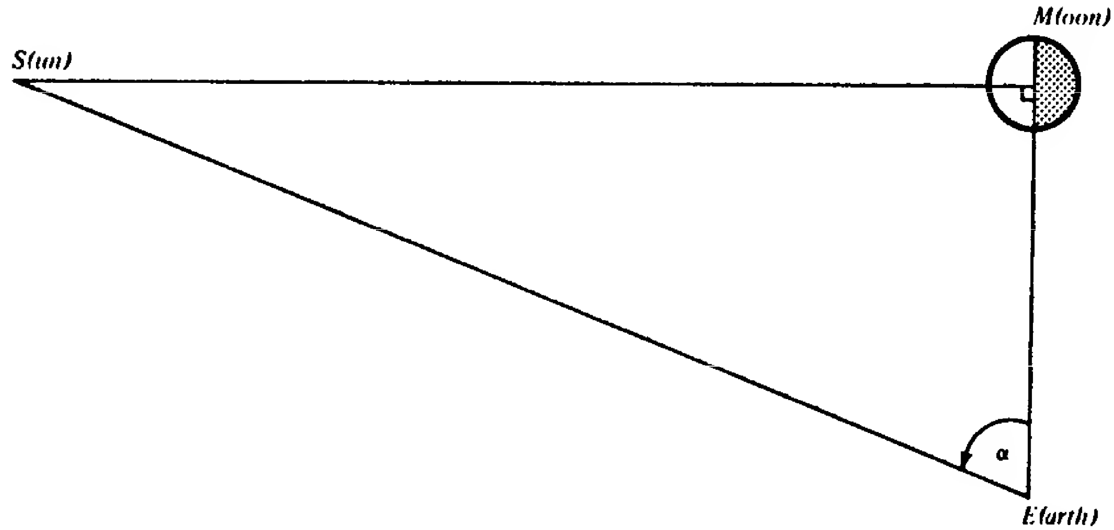

Figure 1.11

Having measured α, Aristarchus had to compute $\cos\alpha$; unlike us he had no tables to refer to. His result was grossly inaccurate for two reasons. α is nearly 90° where a small error is critical. Second, just by looking, one cannot say when precisely the Moon's phase is half-moon; there is a more-or-lessness about the observation. Nevertheless, Aristarchus had a great idea. Since the Sun is vastly more remote than the Moon at half-moon, and since the size of the Sun and Moon as viewed from the Earth remain sensibly constant, it is a safe inference that the Sun is at all times farther from the Earth than the Moon.

1.2.2 Radius of Earth: Eratosthenes

Earlier, in discussing a more interesting question, that of the distance of the Moon from the Earth, we saw that a necessary preliminary is to determine the size of the Earth. So the next important question is: What is the radius of the Earth?

In ascribing radius to the Earth we commit ourselves as to its shape. What shape? Yes, spherical. Is this precisely correct? No, we now know that the Earth is slightly flattened at the poles; it is more nearly an oblate spheroid. But to treat it as a sphere is a good approximation. Good approximations often lead to better ones.

Determination of the Earth's size was Eratosthenes' outstanding achievement. As well as a geographer and astronomer, he was librarian of

the famous library at Alexandria, then the greatest library of the civilized world. He lived from about 280 to 195 B.C., but these dates are problematic. With the subsequent dispersal of this library there is no extent Alexandrian *Who's Who* in which to look him up. Although his dates are in doubt, fortunately his method is not. And so we raise the inevitable question: How did he do it?

The circumstances are as follows. The River Nile flows approximately from south to north, so that the shortest route from the city of Syene (nowadays Aswan) far up the Nile to Alexandria in its delta is a great circle route. That is to say Syene and Alexandria lie (almost) on the same meridian; a circular hoop or belt joining the poles and passing through Syene would also pass through Alexandria. Moreover Egypt is a civilized country, there is a road between Syene and Alexandria, and its length is known. It is 5,000 stadia. See Fig. 1.12. In short, the circular arc AS is 5,000 stadia. If it were known what angle θ at the Earth's center subtended this arc, then it would be known what fraction of the Earth's circumference AS is. The real problem is to determine θ.

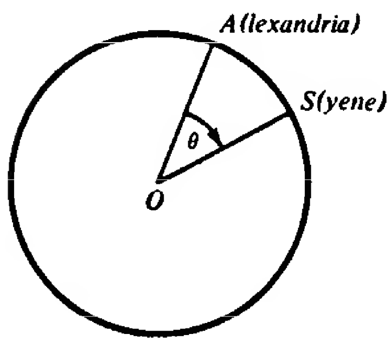

Figure 1.12

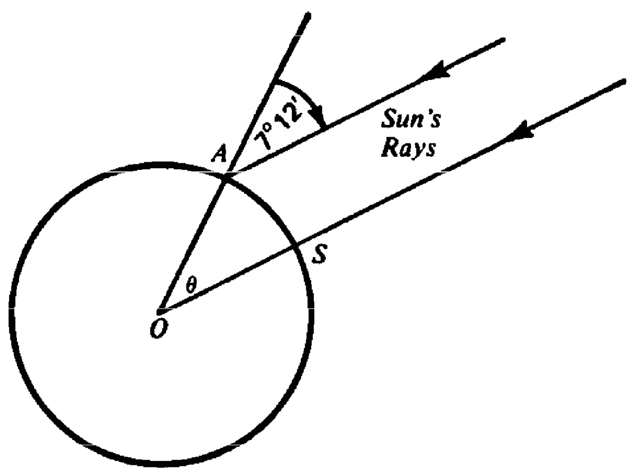

Figure 1.13

Eratosthenes knew that Syene has a very deep well whose waters are touched by sunlight at noon on the longest day of the year, i.e., that there the sun is then directly overhead. So at noon on a midsummer's day he measured the Sun's inclination to the vertical at Alexandria. Of course he needed no watch to tell himself when the Sun's inclination to the vertical was a minimum; to the contrary he used the Sun's minimum inclination to determine noon. He found the angle to be 7°12′. And since the Sun is so remote that its rays are sensibly parallel, the circumstances were as is illustrated by Fig. 1.13. Considering the Sun's rays to be parallel, the

corresponding angles at O and A are equal. We have

$$\theta = 7°12'$$

so that

$$\frac{\theta}{360} = \frac{7°12'}{360°} = \frac{1}{50} \quad \text{of a complete revolution}.$$

Consequently AS is $\frac{1}{50}$th of Earth's circumference. But AS is 5,000 stadia, so that the Earth's circumference is 250,000 stadia and its

$$\text{radius} = \frac{250{,}000}{2\pi} \text{ stadia.}$$

Unfortunately we do not know which of the several stadia used in antiquity is the unit employed by Eratosthenes. A stadium is 600 Greek feet, but the Greeks had several feet; for example, the Attic Stadium is 607 English feet, the Olympic, 630.8 ft. If we take the former, the radius of the Earth becomes

$$\frac{250{,}000}{2\pi} \times \frac{607}{5280} \approx 4{,}600 \text{ miles.}$$

Nowadays the accepted figure for the Earth's equatorial radius is 3,963 miles, for the polar radius 13.5 miles less.

That Eratosthenes' result is inaccurate does not really detract from the greatness of his achievement. It is his method that excites our admiration. Would not a giant measure the Earth by encircling it with his arms to compare its circumference with his span? And what did our little pigmy Eratosthenes do? At Alexandria at noon on a certain midsummer's day long ago, he observed the shadow cast by a little stick and used his protractor. A mere shadow and an idea is the substance that made the pigmy a giant who spanned the Earth.

1.2.3 Rival Cosmologies

How, without a watch, do we know what time it is? Yes, by looking at a sundial. The cast of the Sun's shadow across the dial tells us the time. Despite the fact that a watch has two hands although a sundial has only one "hand", a watch is in effect a sundial. Think about it. The "shadow" or position of the minute hand (read in conjunction with the position of the hour hand) is a substitute for the Sun's shadow. A watch in telling us our time—to be precise, local solar time—indicates our position relative to the Sun. We cannot see in the dark; surely primitive man arose to work

with the rising up and retired to rest with the going down of the Sun. Life was governed by nature's clock.

And how do we measure age? Yes, in years. But, what is a year? The time that elapses before the Earth is again in the same position relative to the Sun. And how do we determine sameness of position? By reference to the framework of the fixed stars. As the position of Earth changes relative to Sun, the days grow longer, then shorter, then longer again. There is a cycle of seasons—of the time to sow and the time to reap. The calendar is our recognition of this periodicity.

Are not our lives regulated by the clock and the calendar? Is not our existence dependent upon the rotation of the Earth relative to the Sun? Without the Sun there would be perpetual night; neither day nor week nor month nor year; neither a time for sowing nor a time for reaping. The fate of all mankind dependent upon the heavens, is it not a natural step to suppose personal destiny to be governed by the stars? Could not greater knowledge of the heavens lead to knowledge of our individual destinies? Although to date astrology has not been a successful application of astronomy, it served a purpose. It gave additional impetus to astronomy; to such solid practical reasons for the study of the stars as the determination of the calendar and a method of navigation, it added its own. Evidence of an especial regard for the wandering stars—the planets—is embedded in our language: Sunday is the day of the Sun; Monday the day of the Moon; Tuesday, via the French, Mardi, the day of Mars; Wednesday, via Mercredi, the day of Mercury; Thursday, via Jeudi, the day of Jupiter; Friday, via Vendredi, the day of Venus; and Saturday, the day of Saturn.

For primitive man nature was a malignant uncertainty. Even for the Greeks, behind every bush and underneath every stone there lurked a god of unpredictable caprice. The paths of the planets gave the comforting assurance of a glimmer of certainty in an uncertain world. These wandering stars as the Greeks called them—in opposition to the fixed stars—appeared to be predestined to follow fixed paths. Planets were observed to reappear in the same position (relative to the fixed stars) at regular intervals. Despite the general fortuitousness of nature, a few events were predictable; their occurrence could be depended upon. In studying the applications of mathematics to astronomy we see the first attempt to discover uniformities in nature. The stars gave man his first glimpse of a great idea—the belief that there are uniformities to be discovered. It is hardly possible to exaggerate the importance of this change of view; the new view is the genesis of science.

Aristotle (384–322 B.C.) argued that a planet must move with uniform motion in a circle. What is his argument? That the planets are necessarily

perfect bodies and therefore spheres and, because perfect, must move with perfect motion, i.e., uniformly in circles. You smile; his contemporaries did not. Aristotle never caused a smile in a thousand years. His dictum persisted without a murmur of contradiction until the Middle Ages. The founder of zoology, of meteorology, of logic had spoken; it was left to lesser men merely to follow in the footsteps of the master and quote his authority.

Circular planetary orbits had been proposed before Aristotle; after Aristotle they were obligatory. The question was: About what center? That the Sun moves around the Earth is a natural impression, and the theory of Hipparchus (160–125 B.C.), developed by Ptolemy (c. 130 B.C.), that all the planets move around the Earth, was generally accepted.

Observation did not precisely fit the theory. So, in the Greek view, if a planet did not move in a circle then its motion must be a combination of *circular* motions. See Fig. 1.14. Here is illustrated a combination of two circular motions. As P moves around the circle of center Q, Q moves around the circle of center C. The path of P (circular or cyclic path relative to Q) is said to be an epicycle. Yet if such a combination did not precisely fit the facts, epicycles of epicycles were tried in later times. See Fig. 1.15. Here is illustrated a combination of three circular motions. As P moves around the epicycle of center Q and Q itself moves around the epicycle of center R, R itself moves around the circle of center C. This point is of importance for the understanding of science; by sufficiently complicating the hypothesis we gain enough flexibility to fit it to our observational data. Fitting the data by an uncomplicated hypothesis is much more interesting.

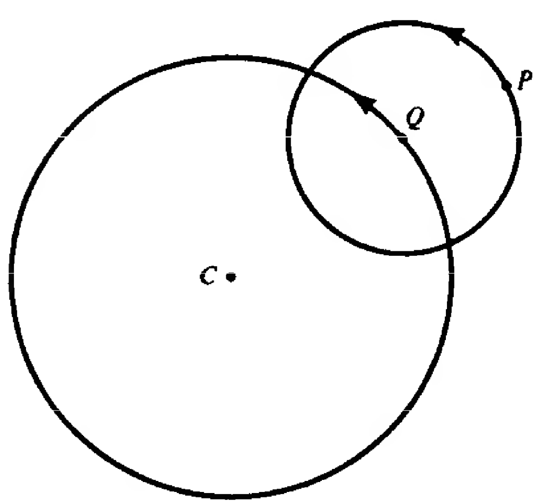

Figure 1.14

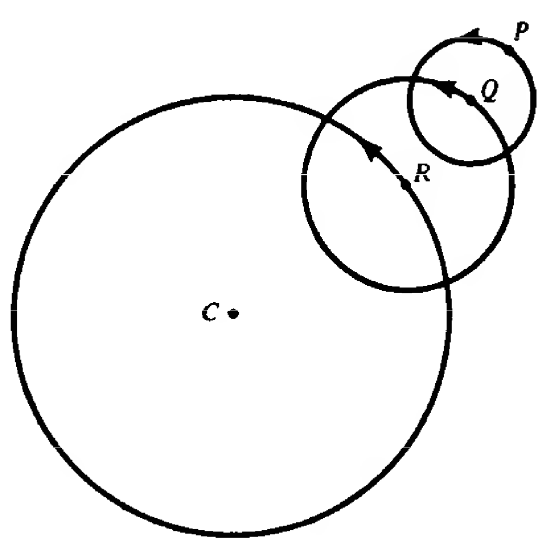

Figure 1.15

We recall a rival theory; that of Aristarchus of Samos. His theory was that the Earth and planets move in circular orbits around the Sun. Although the mass of then available observational data fitted his theory fairly well, it was nevertheless universally rejected; it was rejected by Archimedes (287–212 B.C.), the greatest mathematician, physicist, and inventor of antiquity.

Why was it universally rejected? In part, no doubt, because of Archimedes' rejection. We must remember that pride and prejudice can influence our thinking. Earlier, we asked: How far is the Moon from the Earth? We did not ask: How far is the Earth from the Moon? Both questions must have the same answer, so why the former but not the latter question? When we travel we necessarily start from where we are. Is not the first a more natural formulation? When fogbound in a rowboat is it not more natural to suppose the other fellow's boat drifting past ours than ours drifting past his? Is not an Earth-centered more natural than a Sun-centered theory?

Fig. 1.16 illustrates Ptolemy's universally accepted geocentric (Earth-centered) theory; Fig. 1.17—Aristarchus' universally rejected heliocentric (Sun-centered) theory (in which the Moon orbits the Earth).

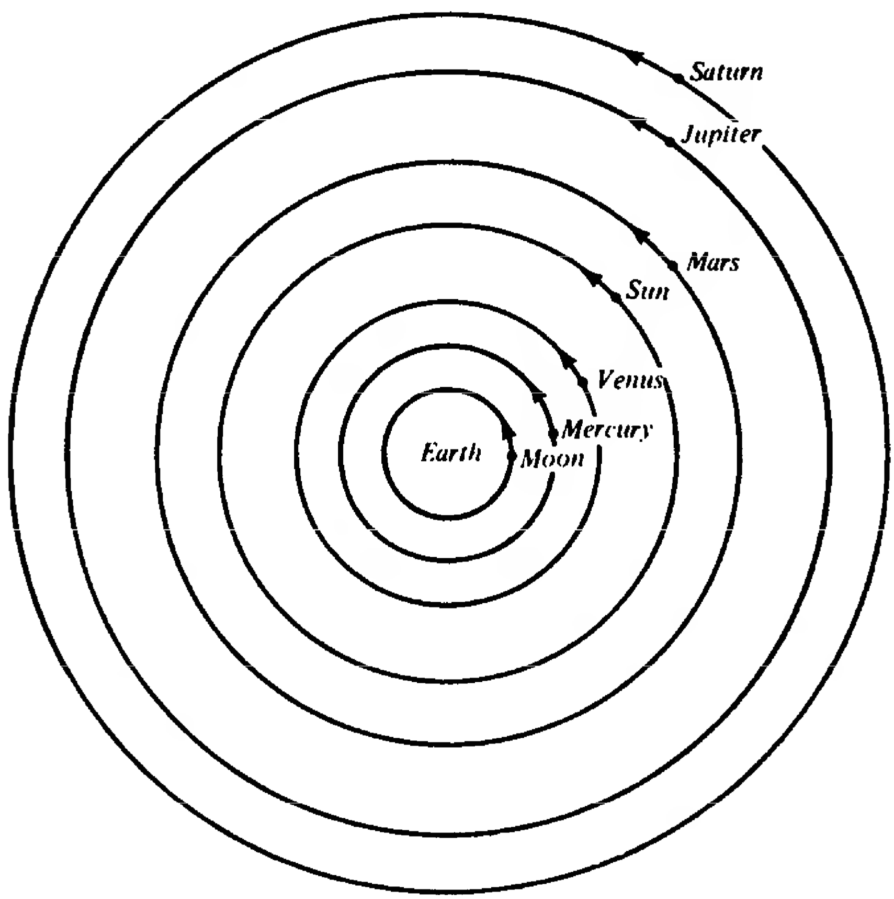

Figure 1.16 Ptolemy's Geocentric System

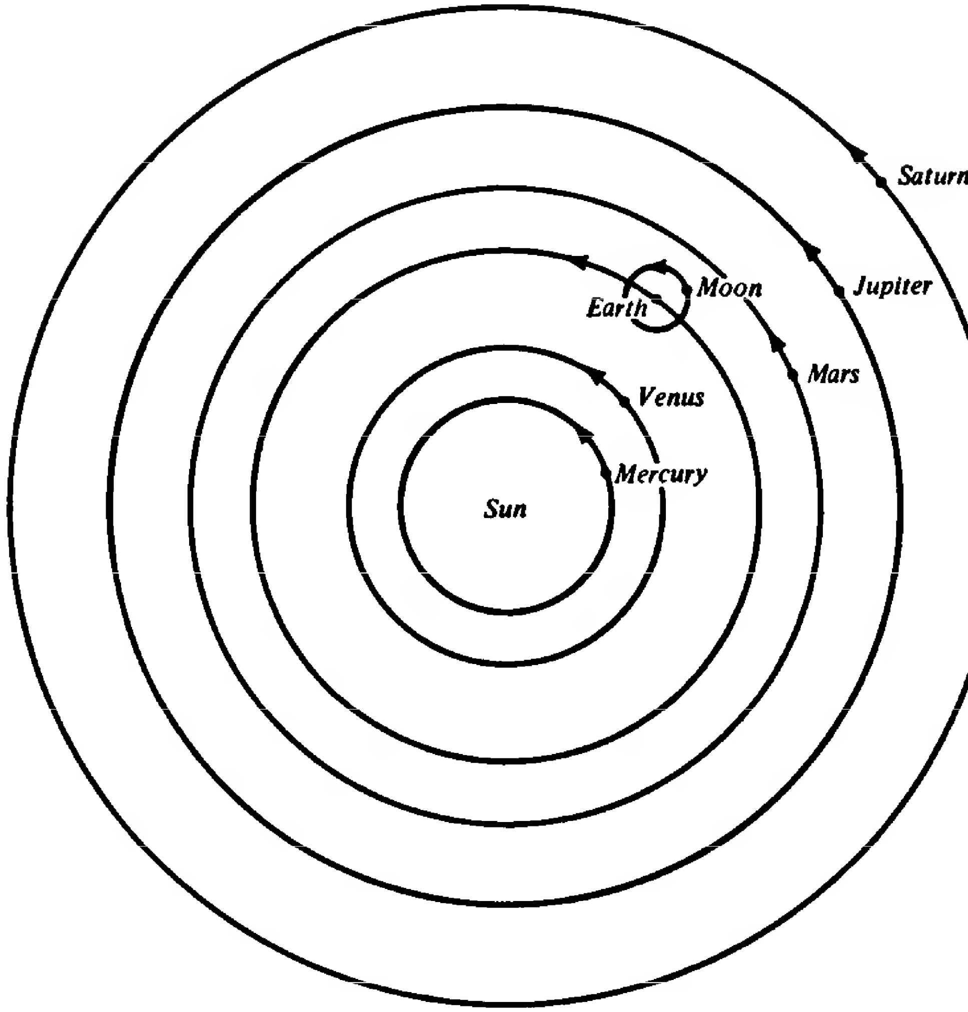

Figure 1.17 Aristarchus' Heliocentric System

Some seventeen centuries later, Aristarchus' heliocentric system was rediscovered by Copernicus (1473–1543). It was still *de rigueur* to quote authority—preferably Aristotle—and as Aristotle could not be quoted in favor of these matters, Copernicus quoted Aristarchus. (Later, however, he deleted this quotation.) As he knew Aristarchus' work it is more correct to say that he redeveloped rather than rediscovered the heliocentric theory. Patiently and pertinaciously, he checked it against a vast accumulation of his own and other astronomers' observations. Although a man of immense intellectual courage as well as energy, he was very careful. Knowing that people do not like to have their old habits of thought, or the habit of not thinking at all, disturbed, he delayed publication of his findings some thirty years until he was upon his death bed. With characteristic caution he did not claim that the Earth and planets do actually move around the Sun; he contented himself with showing that a heliocentric hypothesis works better than a geocentric one: it requires fewer epicycles.

1.2.4 The Orbit of Venus

An earlier theory was propounded by Herakleides who lived in the 4th Century, B.C. He studied under Plato and probably under Aristotle also. His theory is an intermediary between the Ptolemaic and Copernican stand-points. According to Herakleides, Mercury and Venus moved in circles around the Sun, while the Sun itself and all the other planets moved in circles around the Earth.

The bright star often visible at the setting of the Sun is known as the Evening Star; the bright star often visible at the rising of the Sun is known as the Morning Star. Although these names occasion no surprise, surely there was great surprise at the early discovery that the Evening Star and the Morning Star are identical. This star is Venus. Its wanderings, while exhibiting some regularity, were perplexing. Long-term observation showed it eventually to reappear in the same place (relative to the fixed stars); it was at all times relatively close to the Sun, yet sometimes appeared to be moving rapidly in the same direction as the Sun and at others slowly in the opposite direction. But surely perfect bodies, spheres, describe perfect figures, circles, with perfect, uniform, motion. Whatever could be the reason for this apparent discrepancy? A glance at Fig. 1.18 makes the explanation immediately obvious. The trouble with hindsight is that it blinds us to the brilliance of foresight. The explanation is Herakleides'. That we now know that Venus' orbit is not exactly a circle nor its motion precisely uniform detracts nothing from his ingenuity. The astonishing thing is that his hypothesis fits the facts so closely.

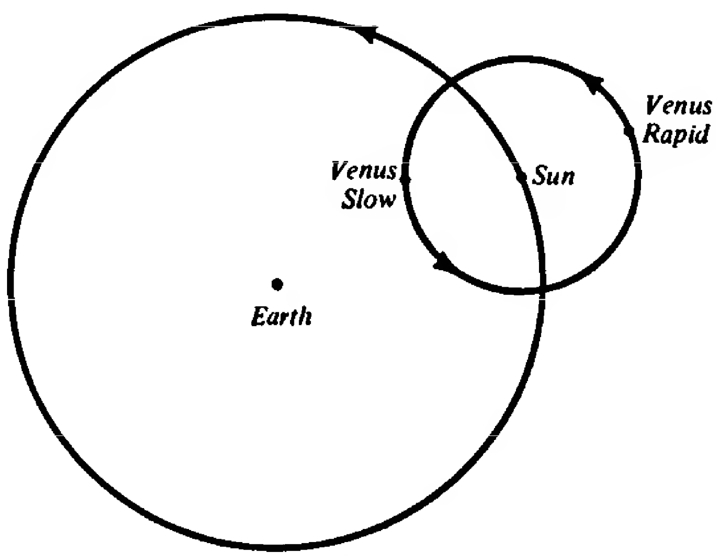

Figure 1.18

The good first approximation accuracy of Herakleides' hypothesis makes it reasonable to ask: What is the radius of Venus' orbit about the

Sun? This question raises another question: How are we to determine this radius? How would you do it? Well, begin with careful study of Fig. 1.19.

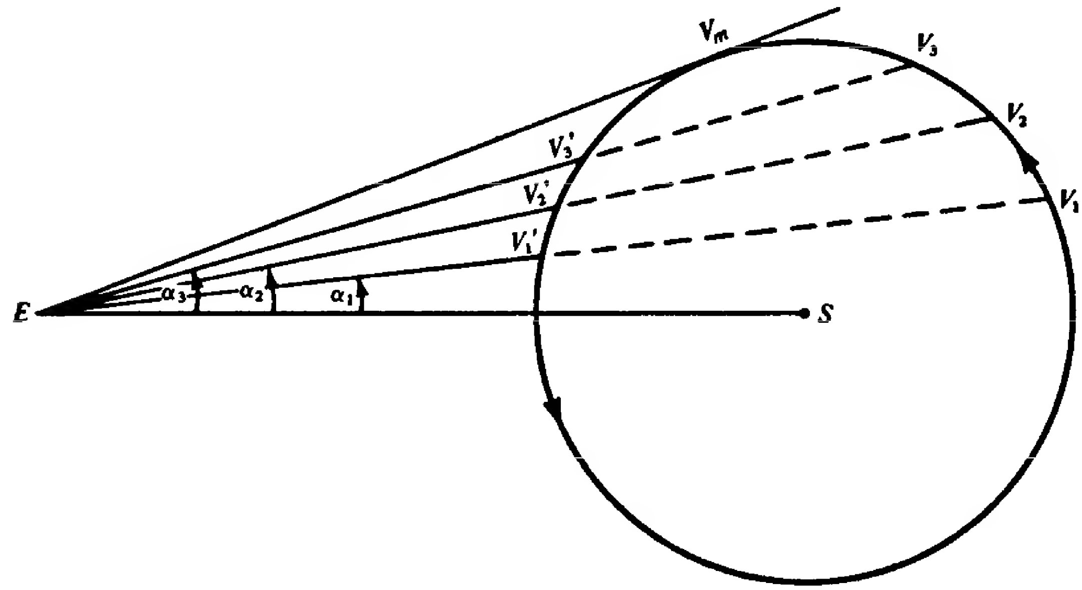

Figure 1.19

Observe that the angle α_m between the straight line joining Venus to the Earth and SE changes as Venus progresses in her orbit. In particular, note that

$$\angle SEV_1 = \angle SEV_1'$$

$$\angle SEV_2 = \angle SEV_2'$$

$$\angle SEV_3 = \angle SEV_3'$$

$$\angle SEV_m = \angle SEV_m.$$

In short, α increases to a maximum when Venus is at V_m (m is for *maximum*) and then decreases. Where is V_m? Consider the successive chords V_1V_1', V_2V_2', V_3V_3'; they are progressively shorter. Obviously EV_m is the limiting position, tangential to the circle. Consequently α is a maximum, say α_m, when $\angle SV_mE$ is a right angle. Consider Fig. 1.20.

Since SV, SV_m are radii of the circular orbit

$$\frac{SV}{SE} = \frac{SV_m}{SE},$$

and since $\angle SV_mE$ of $\triangle SV_mE$ is a right angle

$$\frac{SV_m}{SE} = \sin \alpha_m;$$

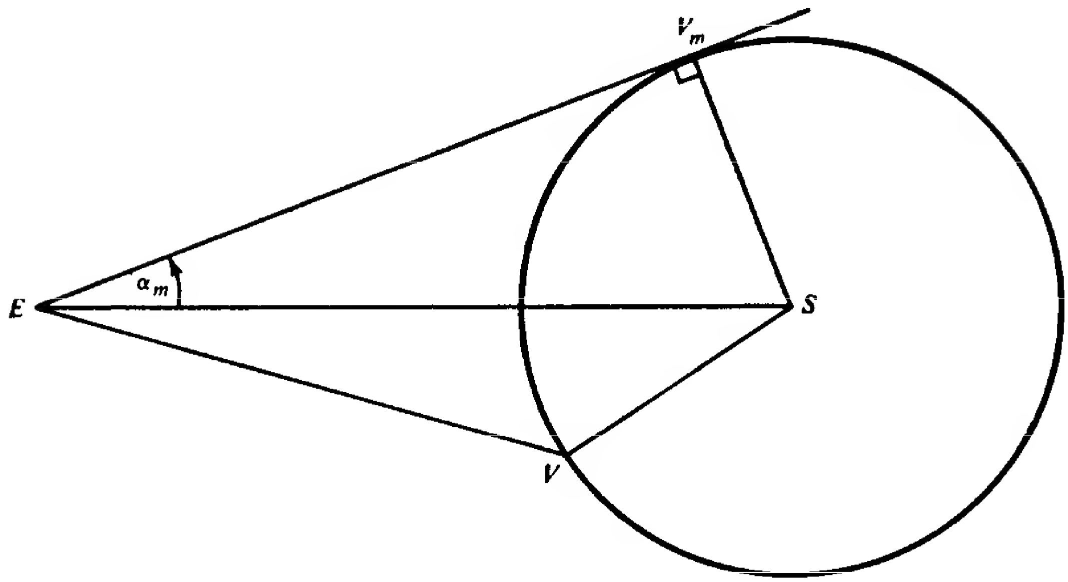

Figure 1.20

consequently

$$\frac{SV}{SE} = \sin \alpha_m.$$

The radius of Venus' orbit is $\sin \alpha_m$ times the distance between Earth and Sun. We do not, of course, have sufficient data to determine the actual distances, only their ratio. Yet is it not surprising what can be done by using the most elementary geometry and trigonometry?

To apply our formula we require the actual numerical value of α_m. How are we to obtain this? By observing Venus when at V_m? But how could we know when Venus is there? To say "when α is a maximum" is to beg the question. The point is that we cannot obtain α_m by a single observation. We cannot have this information for nothing; we must earn it by regularly making measurements at sunset or sunrise day after day. Without a sequence of observations, how could we tell when α ceases to increase and begins to decrease? Advances in science demand tenacity of purpose as well as bright ideas.

1.2.5 Tycho Brahe and Kepler

Tycho Brahe (1546–1601) was a wealthy Danish nobleman who had much land, many serfs, and a quarrelsome disposition. His disposition when he was a young man resulted in a duel, the loss of his nose, the acquisition of a silver substitute, and a marked propensity to shun society. No doubt this sequence of events increased his attraction to astronomy (not a gregarious science). Be this as it may, he had an obsession for

continual and exact observation of the stars, and this obsession is the basis of his fame. No, he did not propound any new theories. Being rich, and also supported by various princes, he was able to have constructed, with utter disregard of the cost, gigantic, well-made instruments that set a new standard of observational accuracy. Nowadays, with accuracy a *sine qua non*, we overlook this vital contribution of Brahe to astronomy and the development of the scientific attitude.

Kepler (1571–1630) was very poor. In his day there were few chairs of astronomy; the patronage of princes was more important. Such patronage was often given for astrology rather than astronomy, and Kepler earned his meager living by the former, thereby enabling himself to study the latter. As he remarks, astrology is the daughter of astronomy, and is it not right that the daughter cares for the mother?

He was a man of genius. His work marks the transition between medieval and modern outlook. For this reason it is called, by Koestler, "The Watershed", in a book of this title. From Kepler the history of thought flows back through a hodge-podge of emerging scientific thought, astrology, mysticism and superstition to Babylonian times, and forward to the modern outlook. His own writings are a mixture of both. Not having the money to buy accurate (and consequently expensive) instruments with which to make his own observations, he finally met Tycho Brahe and inherited his vast accumulation of accurate data, data accurate to a degree that the Greeks would have found incredible. Kepler's ambition was to describe precisely the orbit of Mars. He tried one fruitless combination of epicycles after another. At last, after fourteen laborious unsuccesses, he came to the conclusion that the orbit is neither a circle nor a combination of circles. It must be something else. Kepler's conclusion had astounding novelty; ever since Aristotle's dictum some seventeen centuries earlier, epicyclic motion had been taken as axiomatic. His break with the Aristotelian tradition was the crossing of the watershed.

With industry to match his courage he continued to grind out more and more calculations to test other hypotheses; it was not until nearing the end that the invention of logarithms eased his labors. Ultimately, he hit upon the hypothesis that Mars moves with non-uniform motion in an ellipse with the Sun at one of its foci. Heretical. Utterly heretical. How could the Sun be at one focus rather than the other? How could a planet move with non-uniform motion? How could the universe be so impossibly imperfect? Observation fitted hypothesis like a glove.

The ideal of Euclid's *Elements*, that the theorems are necessarily consequences of the premises, is apt to mislead us into supposing that the development of science has been entirely rational. Nothing could be farther from the truth. Nowhere is irrationality more clearly exhibited

than in the history of astronomy; nowhere in astronomy is prejudice against fact more visible than in the tenaciously conceived notion of perfect bodies in perfect motion.

New theory in astronomy led to a change of world view; a new standpoint, a new civilization. Even in the pre-Sputnik era, some appreciation of these developments was necessarily an ingredient of educated common sense. Surely students will want to know more. A good introductory account is Morris Kline's *Mathematics: A Cultural Approach*. Another is Koestler's large volume *The Sleepwalkers*, of which his above mentioned book, *The Watershed*, is a (large) chapter. This volume is most appropriately titled, for as a sleepwalker with closed eyes finds his way along a roof top, so Aristarchus conjectured the heliocentric system: his facts were few; he knew so little that his eyes were closed—yet he moved with a sure instinct. Later astronomers closed their eyes to facts. Here is a story too fantastic to be fiction, unfolded with spellbinding skill.

1.2.6 The Mars Year

We return to Kepler: How did he discover the precise orbit of Mars? By what observations? And on the basis of what presuppositions? His working hypothesis was that Mars and Earth move in the same plane, each in a circle, with uniform motion around the Sun. We know now as he knew then that this is not precisely correct; their orbits are neither coplanar, nor circles, nor their motions uniform. His hypothesis was just a first approximation; it made, without over-simplification, his problem tractable. An appropriate first approximation is a first step to progressively better approximations.

A consequence of Kepler's hypothesis is that Mars will be in the same *sidereal* position relative to the Sun (that is, as determined against the framework of the fixed stars) at regular intervals. The length of this interval, the time required to complete one orbit around the Sun, is said to be the (sidereal) Mars year. Similarly, of course, the time for Earth to complete one orbit around the Sun is said to be the (sidereal) Earth year. Taking the Earth year as unit of time, Kepler's first task was to determine the Mars year.

Although Mars and Earth move around the Sun in the same direction, they move with different angular velocities. In consequence, Sun and Mars must become, momentarily, exactly opposite each other as seen from the Earth's surface, their longitude then differing by 180 degrees. When this happens Sun and Mars are said to be in *opposition*. See Fig. 1.21.

An opposition is observable with great accuracy. Remember that a full twenty-four hour day is the time interval between two consecutive occurrences of the Sun at its zenith, so that, because of the uniform

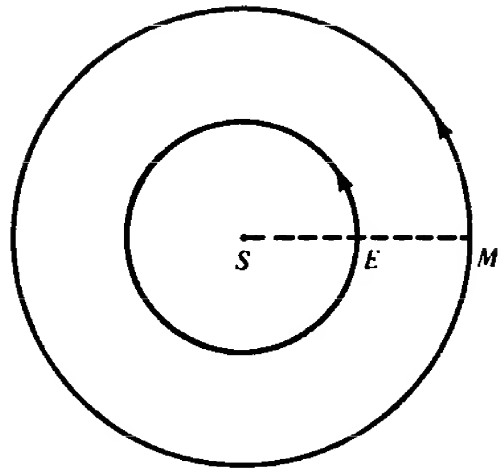

Figure 1.21

rotation of the Earth about its axis, in twelve hours from noon, i.e., at midnight, the Sun is precisely on the opposite side of the Earth. Thus, if at midnight Mars is directly in the meridian, then there is an opposition. The Sun is, so to speak, observed inferentially.

Fig. 1.21 may be considered as the dial of a celestial clock; but the hands are not called hour hand and minute hand: SE is the "Earth" hand and SM the "Mars" hand. We suppose that T_E, the Earth year, the time for the Earth hand to complete a revolution, is known: the Babylonians had determined it with great accuracy. If SM were stationary then obviously the hands would be again collinear after a complete revolution of the Earth hand, i.e., T_E would be the period, P (say), between two consecutive oppositions. If SM were to rotate in the same direction and with the same angular speed as SE, then there would at all times be an opposition; the interval P between consecutive oppositions would be zero. It is equally obvious that if SM were to rotate with the same angular speed but in the opposite direction, then there would be an opposition after SE (and SM) had completed half a revolution, i.e., after time $\frac{1}{2}T_E$. Is it not evident that P, the interval between consecutive coincidences of the hands of our clock, is related to their angular velocities, i.e., that P will depend upon T_E and T_M? Alternatively put, isn't P linked logically between T_E and T_M? The crux of the problem to determine T_M is specification of this relation between T_E, P, and T_M.

Our celestial clock is somewhat peculiar in that the angular speeds of the two hands are not in the proportion 1 : 12 although they are in a constant proportion. Does this make any real difference to the problem? No, of course not.

Let us, with the convenience of brevity, describe the position of the

hands at the opposition of Fig. 1.21 as on the *initial line*. What happens subsequently? Because the Earth hand rotates faster it necessarily completes a revolution before the Mars hand does. Thus, when SE arrives at the initial line, SM has but partially completed a revolution. See Fig. 1.22.

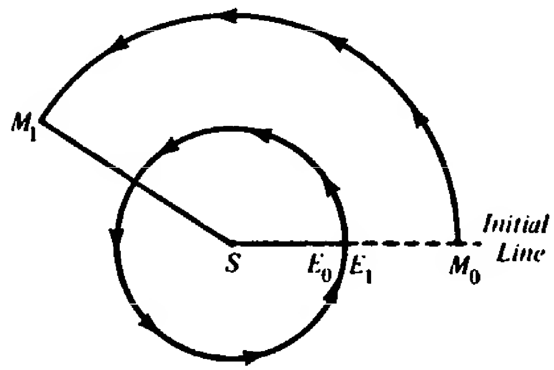

Figure 1.22

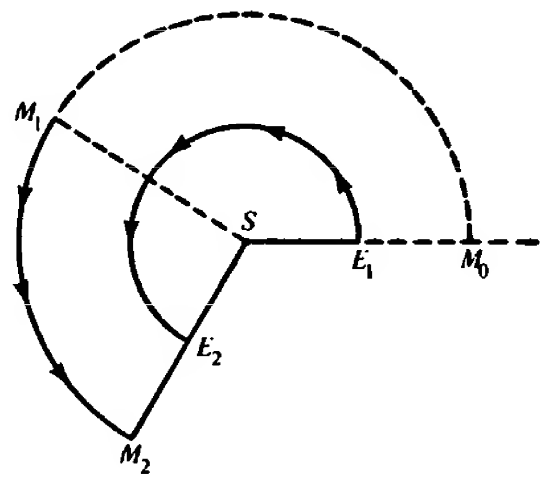

Figure 1.23

Ignore for the moment that the situation of Fig. 1.22 derives from that of Fig. 1.21. Concentrate on what follows on from Fig. 1.22. The ensuing situation is analogous to that of a handicap race: E, on the starting (initial) line, is handicapped by M starting way ahead at M_1. But, the angular velocity of SE being greater than that of SM, E must sooner or later catch up with M. Suppose this to occur when SE has rotated through an angle α (measured, of course, from the starting line). So at the end of the race the circumstances are as illustrated by Fig. 1.23. While SM has rotated from SM_1 to SM_2, SE has rotated through an angle α from the initial line to SE_2. Now recall Fig. 1.21. During the interval between one opposition and the next, SM has rotated from the initial line to SM_2; i.e., has turned through an angle α. And remember that SE completed a revolution before the start of the handicap race. We conclude that if in the period P between two consecutive oppositions, SM rotates through α, then SE rotates through $360° + \alpha$.

It pays to look back. Isn't this conclusion immediately obvious to hindsight? We now know the right way of looking at the problem: immediately after an opposition the Earth hand SE forges ahead and so will have to rotate 360° more than the Mars hand SM in order to catch up with it.

The rest is plain sailing. Tabulating our data for Mars and Earth, we

have:

	Time interval	Angle rotated through in the time interval
Mars	P T_M	α 360
Earth	P T_E	$360 + \alpha$ 360

But, for a (clock hand) rotation with uniform angular velocity the time of rotation is directly proportional to the angle of rotation. Therefore, from the Mars data, we get

$$\frac{P}{T_M} = \frac{\alpha}{360}, \tag{1}$$

and from the Earth data, we get

$$\frac{P}{T_E} = \frac{360 + \alpha}{360}. \tag{2}$$

To link T_M to T_E by P it remains merely to eliminate α. From (2),

$$\frac{P}{T_E} = 1 + \frac{\alpha}{360}.$$

Hence, by (1)

$$\frac{P}{T_E} = 1 + \frac{P}{T_M}.$$

We have established the explicit expression of the relation between T_M, P, and T_E. The latter being known, it remains merely to measure P in order to compute T_M. P had been measured by the Greeks; Kepler computed T_M.

1.2.7 The Orbit of Mars

Recalling that Kepler's ambition was to determine precisely the orbit of Mars, the alert reader will ask: How is the determination of T_M instrumental to this end?

Consider Fig. 1.24. When Mars is at M, collinear with the Sun S and

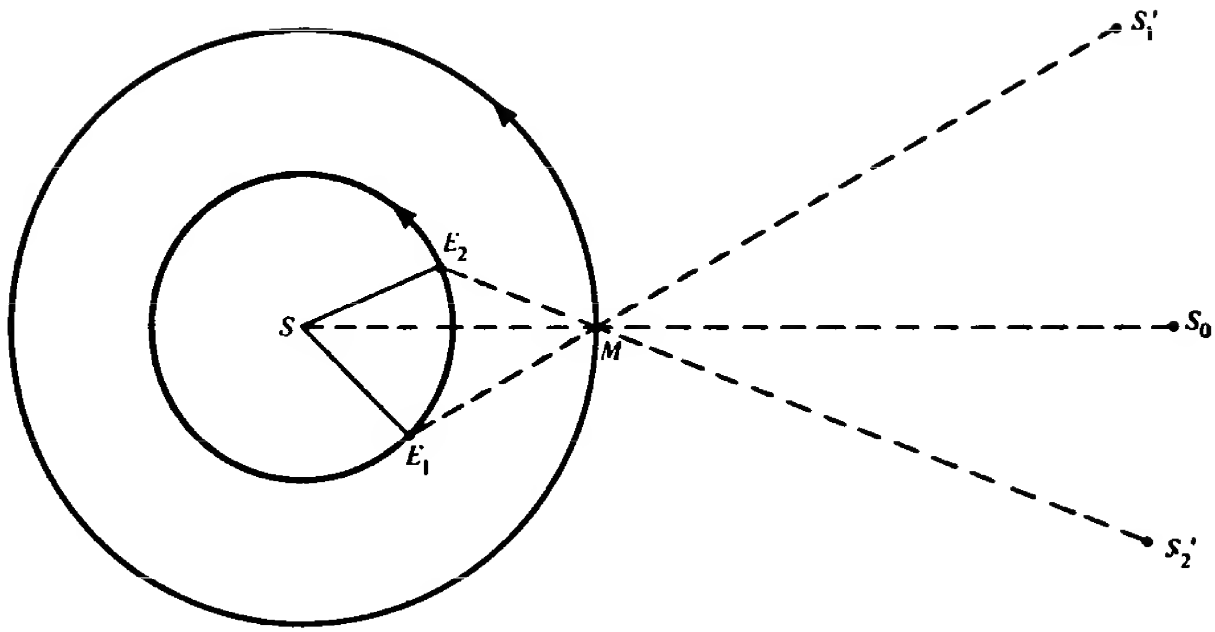

Figure 1.24

a fixed star S_0, suppose the Earth to be at E_1. Precisely T_M later Mars will have completed an orbit around the Sun and again be at M, but since the Earth hand of our celestial clock rotates faster than the Mars hand, the Earth will have rotated more than a complete revolution and be at E_2. Although Mars is again in its initial position relative to the fixed stars as viewed from the Sun (i.e., collinear with S_0), it is in a different position relative to the fixed stars as viewed from the Earth. Initially Mars is collinear with Earth and S_1'; T_M later, with Earth and S_2'. Yet despite Mars at M *appearing* against the framework of the fixed stars as in different positions when viewed from the Earth at E_1 and E_2, since the Mars year is T_M, we know *inferentially* that it is in the same position.

We may infer much more. T_E being known, the angular velocity of our celestial Earth hand is known, so that the angle turned in time T_M can be computed: we can determine $\angle E_1SE_2$. And taking the radius of the Earth's orbit as known, the length of base E_1E_2 and base angles of isosceles triangle E_1SE_2 are determinate.

What else do we need to compute SM? What are the easiest things to measure accurately? Yes, angles. Now consider Fig. 1.25. The position of the Earth relative to Sun and fixed stars had been given careful study from Babylonian times; Tycho Brahe had made most exact observations. This data enabled Kepler to determine the fixed star S_1 collinear or most nearly collinear with Sun and Earth when, for example, the Earth was at E_1. We have already remarked that although the Sun is not visible to the terrestial astronomer at E_1 when observing S_1, it is nonetheless "observable inferentially". Thus Kepler was able to measure $\angle S_1E_1M$ (where

S_1E_1 produced passes through S) when the Earth was at E_1 and likewise $\angle S_2E_2M$ (where S_2E_2 produced passes through S) when, later, the Earth was at E_2. (Do not confuse the fixed star S_1 of Fig. 1.25 with S_1' of Fig. 1.24: S_1 is collinear with E_1 and S; S_1' is collinear with E_1 and M. Nor is S_2 to be identified with S_2'.)

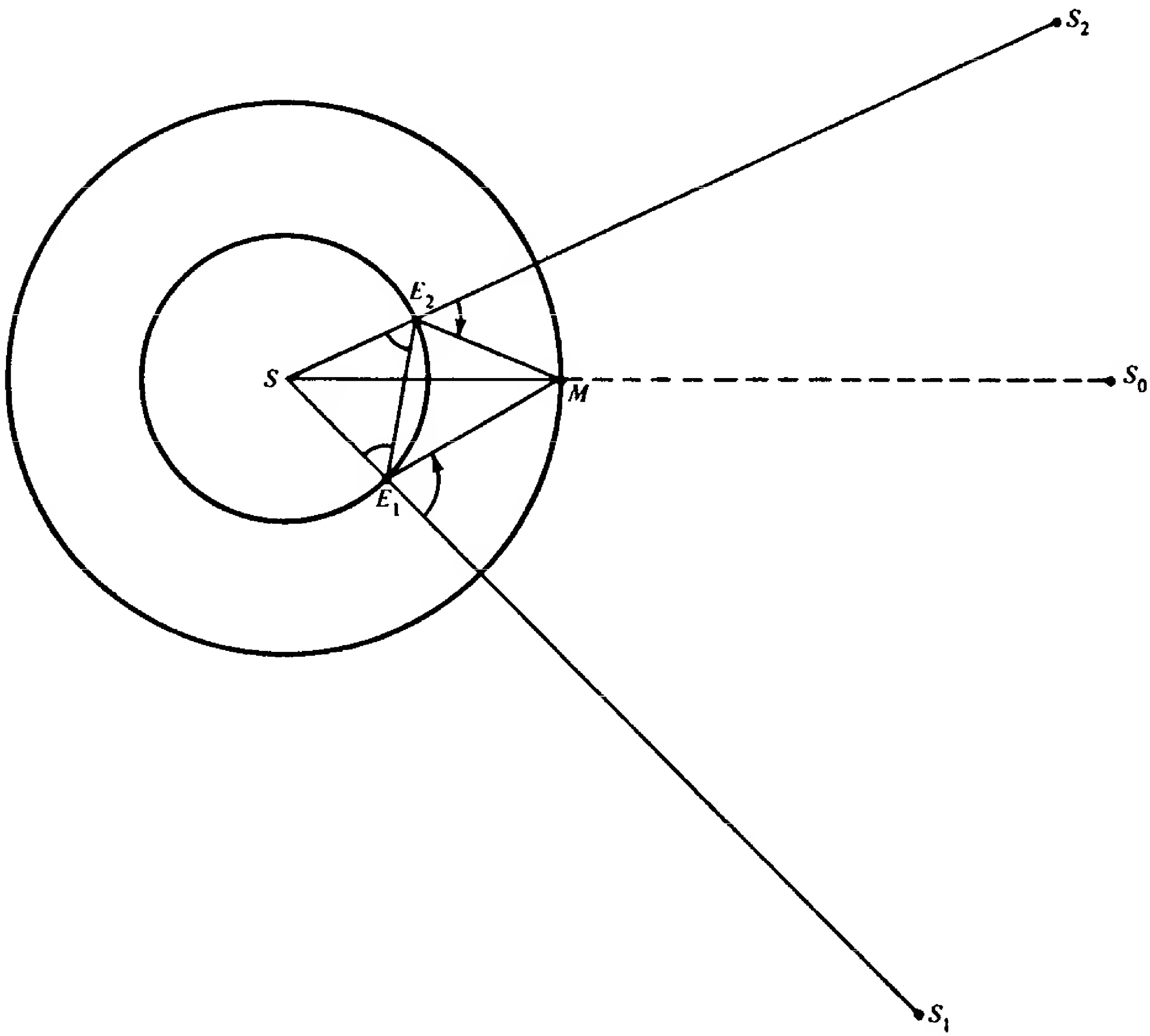

Figure 1.25

What use can be made of this additional data? $\angle E_2E_1M$ is the supplement of the sum of the known angles, $\angle SE_1E_2$, $\angle S_1E_1M$, and therefore determinate. $\angle E_1E_2M$ is similarly determinate. Thus in $\triangle E_1E_2M$ two angles and a side (E_1E_2) are known, so that by using the sine law E_2M is computed. But $\angle SE_2E_1$ and $\angle E_1E_2M$ are both known, so that in $\triangle SE_2M$ we know the angle SE_2M as well as the sides containing it; by using the cosine law, SM is computed.

Remember that Kepler's working hypothesis includes the supposition that Mars repeats its orbit with regularity: no matter what its orbital

position at a specified time, it will again be in that position after an interval T_M. So the above method is applicable to computation of the length of the Mars hand of our celestial clock in any position. In this way T_M was instrumental to Kepler's determination of Mars' orbit around the Sun.

Having computed many radius vectors of Mars' orbit, Kepler with energy to equal his enthusiasm set about fitting theory to fact. His inheritance of Tycho Brahe's observations gave him data with an accuracy unknown to the Greeks—and consequently made his task all the more difficult. Finally, at his fourteenth attempt, the theoretical orbit consequent upon his hypothetical epicycles closely approximated to the factual orbit: there was a discrepancy of merely eight minutes of arc, an accuracy unknown to the Greeks. But closeness of fit which would have been more than good enough to satisfy the Greeks was rejected out of hand by Kepler. And with it he rejected the notion of cycle and epicycle, bag and baggage. He was sick with the wearisome repugnance of epicycle piled upon epicycle; the dogma of perfect motion had become a celestial nightmare. His final hypothesis was that Mars moved in an ellipse with the Sun at one focus: it worked.

This, in rough outline, is how Kepler discovered the first mathematical law of astronomy. Unfettered from the dogma that the planets move in perfect figures, i.e., circles, it was an easy step to reject also the fiction that they move with uniform velocity. The hands of our celestial clock rotate with variable speed. Tycho Brahe's observations afforded ample evidence. Indeed, it was known to the Greeks that the nearer the Earth is to the Sun the faster it moves; yet it took the insight of genius to discover the law. See Fig. 1.26.

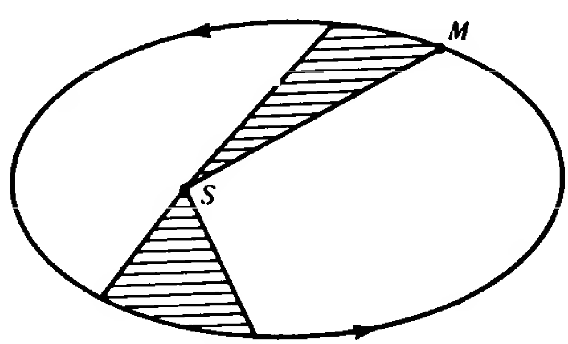

Figure 1.26

Kepler found likewise that the farther Mars was from the Sun, the slower it moved; the nearer, the faster. Eventually he discovered the law which fits the facts. Mars moves in its orbit so that the radius vector SM sweeps out equal areas in equal times.

By analogy Kepler extended his two laws for Mars to the other planets. The available data fitted.

Many years later he discovered a third law. We recall that the planets in the order of their distances from the Sun are Mercury (nearest), Venus, Earth, Mars, Jupiter, and Saturn (most distant). Also it is a fact that the farther a planet from the Sun, the longer it takes to complete its orbit. Kepler first supposed that T, the planet's year, is proportional to R, its mean radius about the Sun. He quickly found that T increases faster than direct proportion; to double R more than doubles T. The law is hidden; eventually Kepler found it: The square of T is proportional to the cube of R.

Kepler's published work is a hodge-podge of astronomy, astrology, geometry, theology, and a miscellany of oddments: he sat astride the watershed. Yet it is intensely interesting, for unlike Galileo and Newton he did not try to cover his traces. His conjectures, failures, successes, errors, insights, fallacies, obsessions, are all revealed with disarming frankness. No other man of genius has been so open about his wild goose chases. But Kepler's work is so full of competing ideas that it remained for Newton to separate the wheat from the chaff, to discern the importance of what Kepler did not himself fully appreciate—his three laws.

1.2.8 A Word to the Perceptive Reader

What is the primary importance of Kepler's work for the student of mathematics and its role in science? First that there are applications of trigonometry on the grand scale. Trigonometry, as we have seen, made computation of Mars' radius vector possible. What could even a Kepler have done without mathematics?

Second, we see the role of what, usually ill-described as "trial and error", is better called *successive approximation*. Kepler, we recall, starting from the working hypothesis of uniform circular motion, determined the Mars year T_M only to conclude finally that Mars' motion is neither circular nor uniform, but elliptic and non-uniform.

Doesn't this appear paradoxical? The initial hypothesis that Earth and Mars have uniform circular motion is erroneous, yet a good approximation to the truth. Note that the calculation of the period is not invalidated by the orbits of Earth and Mars being non-circular: the coincidence of our celestial clock's hands is independent of variations in their length and dependent only on the uniformity of their rates of rotation. Also, as luck would have it, variation in the Earth hand's angular velocity is less than that of the Mars hand, so that a good approximation to the Earth's orbit suffices to show that a similar assumption for Mars' orbit is unacceptable.

More accurate observation of the Earth's orbit leads to more accurate determination of Mars'.

1.2.9 Newton's Problem of a Comet's Path

We conclude this section with a problem. Calculus is not necessary, but you will need your trigonometry. Newton, in addition to his monumental *Principia* took the trouble to write a book on what we now call high school algebra. And what is the main point of Newton's algebra text? The same as Descartes': to solve word problems—thereby demanding, among other things, the full comprehension necessary to translation of problems from prose into mathematics. Newton's problem in good old-fashioned English is: "To determine the position of a comet's course, that moves uniformly in a right line, from three observations." Fig. 1.27 illustrates the problem. Newton knew perfectly well that a comet does not move uniformly and does not move in a straight line. What is the path of a comet? Yes, an ellipse. But, don't you see, a straight line is a first approximation? Here is the first step in a sequence of successive approximations. What is observable? O stands for *Observer*. O observes the comet at A, at B, and at C and notes in each position the star with which it coincides or most nearly coincides. The angles subtended at O by these fixed stars are measured, i.e., ω and ω' are known. Also O observes *when* the comet is at A, at B, and at C, so that the times t and t' for the comet to pass from A to B and A to C are known. In

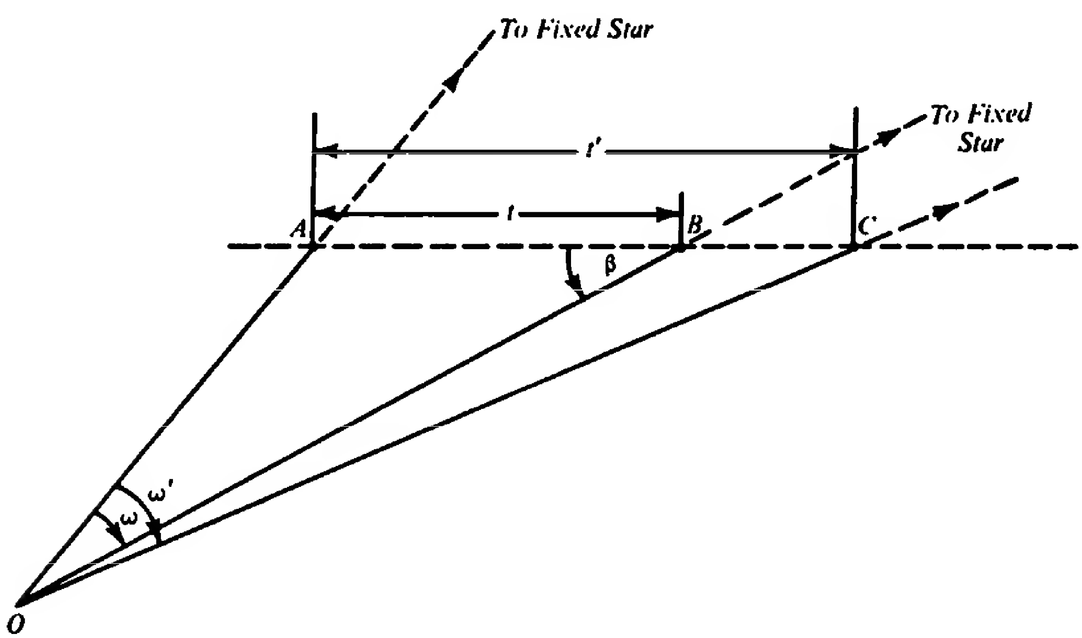

Figure 1.27

short, given ω, ω', t, t', and that the comet's motion is uniform, we are required to find the direction of ABC. This is most conveniently determined by finding β. We conclude with one hint: to find β we must first find a trigonometric function of β—the nicest is the cotangent.*

SECTION 3. SUCCESSIVE APPROXIMATION

We begin by repeating a point: it is important. Kepler started from the assumption that Earth and Mars move with precisely uniform circular motion around the Sun as center, to arrive finally at the conclusion that Mars moves neither in a circle, nor uniformly, nor is the Sun the center of its orbit. To the uninitiated, his argument, like the orbit of Mars itself, appears to be circular. But scientists habitually argue this way: from a working hypothesis given by proposition ρ we are led to the conclusion "no, not ρ". This procedure is often described as the method of *false position*. From a "false" (inaccurate) start we proceed to a "true" (accurate) finish: beginning with what is only approximately correct, we reach by successive diminution of error, if not a dead accurate result, a much closer approximation.

The method is well illustrated by the way we look up a word: say CONFIANCE, in a French dictionary. We open the dictionary at where we estimate the word to be. If the page does not contain words beginning with C then we have made a poor estimate; we have judged falsely the position of the word; we have made a false start. But a poor estimate can be a step in arriving at a better one; a false position can lead to a truer (more accurate) one. Suppose our first estimate turned out to be a page of words beginning with B; we estimate that we must turn on five or six pages. Doing this we turn up, say, words beginning with CA. We have arrived at an improved position. But we want C to be followed by O, not by A; we have found the word correct to the first, but not the second, letter. If our next estimate gives us CO . . . we have found the word correct to at least two letters: if incorrect and faced with CL we turn forward; if faced with CZ we turn back. Well, you know how to use a dictionary efficiently—but did you appreciate that in so using we employ the method of *false position*, or more aptly put, *successive approximation*? Isn't the underlying general idea like that of successively computing a square root to one, then two, three, four, . . . places of decimals?

Full appreciation of a mathematical method cannot be had by merely

*See *Mathematical Discovery*, Pólya, Vol. 1, p. 54, problem 2.63.

talking about it, only by intelligently using it. So, let us apply the method and see the idea in action.

1.3.1 First Application

Consider the following problem, which is a good one for an eight-year-old. If the price of a loaf is one quarter (25 cents) and half a loaf, what is the price of a loaf? Can you do it in your head? Try. Let x cents be the price of a loaf, then

$$x = 25 + \frac{x}{2}$$

so that

$$x = 50.$$

A loaf costs half a dollar.

There is for our purposes a much more instructive way of doing it. To prevent ourselves sinking too deeply into the particularity of the problem, we generalize by taking a cents instead of 25 cents. This is an enormous advance which enables us to deal *en bloc* with a whole family of problems; our problem, its brothers, sisters, cousins and aunts. Generalized, the problem is: find x, given that

$$x = a + \frac{x}{2}. \tag{1}$$

Obviously, the solution is $x = 2a$; yet a person with the outlook of a practical engineer might be enticed to tackle our problem in the following complicated, but most ingenious, way: $x/2$ is less than x, so let us neglect it, and our initial approximation x_0 is

$$x_0 = a. \tag{2}$$

Obviously this approximation is too small, but it is only a first trial. Surely we can do better than this? What happens when we substitute (2) in the right-hand side of (1)? Let's find out. Our new approximation x_1 (conveniently called "the first") is

$$x_1 = a + \frac{x_0}{2} = a + \frac{a}{2}.$$

This is better, so let's repeat the procedure, i.e., consider a second approximation x_2 such that

$$x_2 = a + \frac{\text{the preceding approximation}}{2}$$

i.e.,

$$x_2 = a + \frac{x_1}{2} = a + \frac{1}{2}\left(a + \frac{a}{2}\right) = a + \frac{a}{2} + \frac{a}{4}.$$

Better still. Nothing succeeds like success. What is x_3?

$$x_3 = a + \frac{x_2}{2} = a + \frac{1}{2}\left(a + \frac{a}{2} + \frac{a}{4}\right) = a + \frac{a}{2} + \frac{a}{4} + \frac{a}{8}.$$

Satisfy yourself that

$$x_4 = a + \frac{a}{2} + \frac{a}{4} + \frac{a}{8} + \frac{a}{16}.$$

We can repeat this procedure again and again. Although we will never reach the true value, we can come closer and closer to it. This is readily seen in the following way. Take a number line in which numbers appear as distances, so that a is represented by an abscissa. See Fig. 1.28.

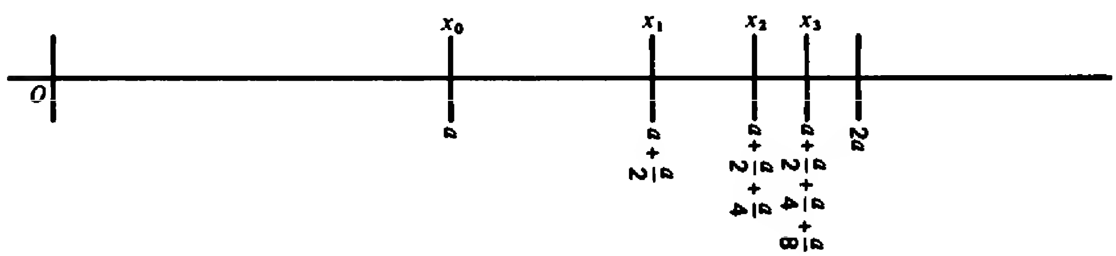

Figure 1.28

Observe that x_0 (i.e., a) is a halfway between 0 and $2a$, x_1 (i.e., $a + a/2$) is halfway between x_0 and $2a$. x_2 (i.e., $a + a/2 + a/4$) is halfway between x_1 and $2a$, and x_3 (i.e., $a + a/2 + a/4 + a/8$) is halfway between x_2 and $2a$. In other words, x_1 is obtained by adding half the difference between x_0 and $2a$, x_2 is obtained by adding half the difference between x_1 and $2a$, and so on. In making the nth approximation x_n we take half the preceding difference (i.e., between x_{n-1} and $2a$). Since we *take* half we *leave* half; since we leave half there is always half left. Thus no matter how many successive steps we take we will never get the exact solution to (1), yet every step must give a better approximation than its predecessor. Observe that x_1 is $a/2$ short of $2a$, x_2 is $a/4$ short of $2a$, x_3 is $a/8$ short, x_4 is $a/16$ short. But these denominators, 2, 4, 8, 16, . . . are powers of 2. Putting this

explicitly, we have

$$\begin{aligned} x_1 &= a + \frac{a}{2} = 2a - \frac{a}{2} &&= 2a - \frac{a}{2} \\ x_2 &= a + \frac{a}{2} + \frac{a}{2^2} = 2a - \frac{a}{2^2} &&= 2a - \frac{a}{2^2} \\ x_3 &= a + \frac{a}{2} + \frac{a}{2^2} + \frac{a}{2^3} &&= 2a - \frac{a}{2^3} \\ x_4 &= a + \frac{a}{2} + \frac{a}{2^2} + \frac{a}{2^3} + \frac{a}{2^4} &&= 2a - \frac{a}{2^4}. \end{aligned}$$

We conclude that

$$x_n = a + \frac{a}{2} + \frac{a}{2^2} + \frac{a}{2^3} + \ldots + \frac{a}{2^n} = 2a - \frac{a}{2^n}. \tag{3}$$

Our algebra confirms our geometry.

This result invites generalization. What is the pattern exhibited by the sequence of terms of x_n? Each term (except of course the first) is $\frac{1}{2}$ of the one before it. But our sequence would still exemplify this pattern if instead of the ratio of a term to its predecessor being $\frac{1}{2}$, it were $\frac{1}{3}$, or $\frac{4}{5}$, or $\frac{11}{9}$. Generalizing, let the ratio of any term to its predecessor be r. The first term is a; what is the second? Yes, ar^1, and the third term is ar^2. What is the nth term? There are $n - 1$ terms after the first; and, with each, another factor r is introduced, so that the nth term is ar^{n-1}. Let S_n be the sum of n terms, then

$$S_n = a + ar + ar^2 + ar^3 + \ldots + ar^{n-2} + ar^{n-1}. \tag{4}$$

Any sequence of this pattern—in which each term is in the same ratio to its predecessor—is said to be a geometrical progression.

Since any term is obtainable by multiplying its predecessor by r, it follows that to multiply each of the first n terms by r is to obtain the sequence of n terms which begins with the 2nd and ends with the $(n + 1)$th. We have

$$\begin{aligned} S_n &= a + (ar + ar^2 + ar^3 + \ldots + ar^{n-2} + ar^{n-1}) \\ r \cdot S_n &= (ar + ar^2 + ar^3 + \ldots + ar^{n-2} + ar^{n-1}) + ar^n. \end{aligned}$$

Subtracting,

$$(1 - r) \cdot S_n = a + (0 + 0 + 0 + \ldots + 0 + 0) - ar^n,$$

so that

$$S_n = a \cdot \frac{1 - r^n}{1 - r}. \tag{5}$$

Does (5) check against (3)? If in (4) we put $r = \frac{1}{2}$, we have

$$S_n = a + \frac{a}{2} + \frac{a}{2^2} + \ldots + \frac{a}{2^{n-1}}.$$

Comparing this with (3) we note that x_n has an extra term, i.e., $(n + 1)$ terms: to make (4) and (5) applicable we must write $n + 1$ for n. Doing this in (4) and (5), (with $r = \frac{1}{2}$, of course) and equating them, we have

$$\begin{aligned} S_{n+1} &= a + \frac{a}{2} + \frac{a}{2^2} + \ldots + \frac{a}{2^{n-1}} + \frac{a}{2^n} = a \cdot \frac{1 - \left(\frac{1}{2}\right)^{n+1}}{1 - \frac{1}{2}} \\ &= \frac{a}{\frac{1}{2}}\left(1 - \frac{1}{2^{n+1}}\right) = 2a\left(1 - \frac{1}{2^{n+1}}\right) \\ &= 2a - \frac{a}{2^n}. \end{aligned}$$

It checks. Solution of our little problem about the cost of a loaf, by the engineer's method of successive approximation, has led to discovery of the formula for the sum of a geometric progression. Oh yes, doubtlessly it has been discovered thousands of times, but this makes the result no less a natural outcome of our deliberations.

1.3.2 Extraction of Square Roots

Let us consider another example of successive approximation: $\sqrt{2}$. Perhaps you can quote it from memory correct to a few decimal places, and probably you learned the "standard textbook method" of square root extraction in school. You know the method I mean: start at the decimal point and pair off the digits before and after it, then from the left-most number, i.e., the number denoted by the digit or pair of digits of the first "pair," subtract the greatest perfect square not exceeding it, and . . . . Well, you know how it goes. Of course you can use this method if you

insist—I had it inflicted on me almost seventy years ago. I didn't understand the reason for it, so I hated it. I still do. But you can use a different method; more interesting, because it is immediately intelligible; more useful, because it has an important generalization. By hand it is a little slower to use than the "standard textbook method," with a desk calculator, quicker. You have to make up your mind whether you prefer rapid computation or a method which leads somewhere.

To avoid duplication of material, the reader is referred to the "divide and average" method of finding square roots. See, for example, *High School Mathematics*, Unit 3, pp. 121–130, University of Illinois Press, Urbana, 1960; *First Course in Algebra*, pp. 304–307, Yale University Press, New Haven, 1960.

SECTION 4. NEWTON'S METHOD OF SUCCESSIVE APPROXIMATION

Successive approximation is an important mathematical method; it is the very essence of science. Although almost invariably in science we must begin with what is only an approximation to the truth, we need not rest content with it. A crude approximation can be made to lead to a less crude approximation; a good approximation to a better one. That the notion of successive approximation is a key to more exact knowledge makes it a worthwhile study.

1.4.1 The General Method of Newton

Newton devised a general method to find the roots of an equation, that is, to find the values of x such that $f(x) = 0$. First of all, to get some idea where the roots lie, we sketch the functional relation $y = f(x)$. Suppose that part of our curve is as illustrated by Fig. 1.29. Note that at P where the curve crosses the x-axis, the y value or ordinate is zero, so that $f(x) = 0$. Thus, in other words, the x value or abscissa of P is a root of the equation. The problem is how to get successively better estimates of this abscissa. It is, of course, sufficient to consider one root of the equation, for once we are clear as to the method, *mutatis mutandis*, other roots can be found.

Our graph enables us to get started, to make an initial estimate. Suppose our estimate of the abscissa of P to be x_0. (This is a convenient notation, for the subscript O may be construed mnemonically as *Original* approximation.) Is x_0 dead accurate? Just possibly, but generally we cannot expect to be so lucky. We test it. Substituting x_0 for x, we find

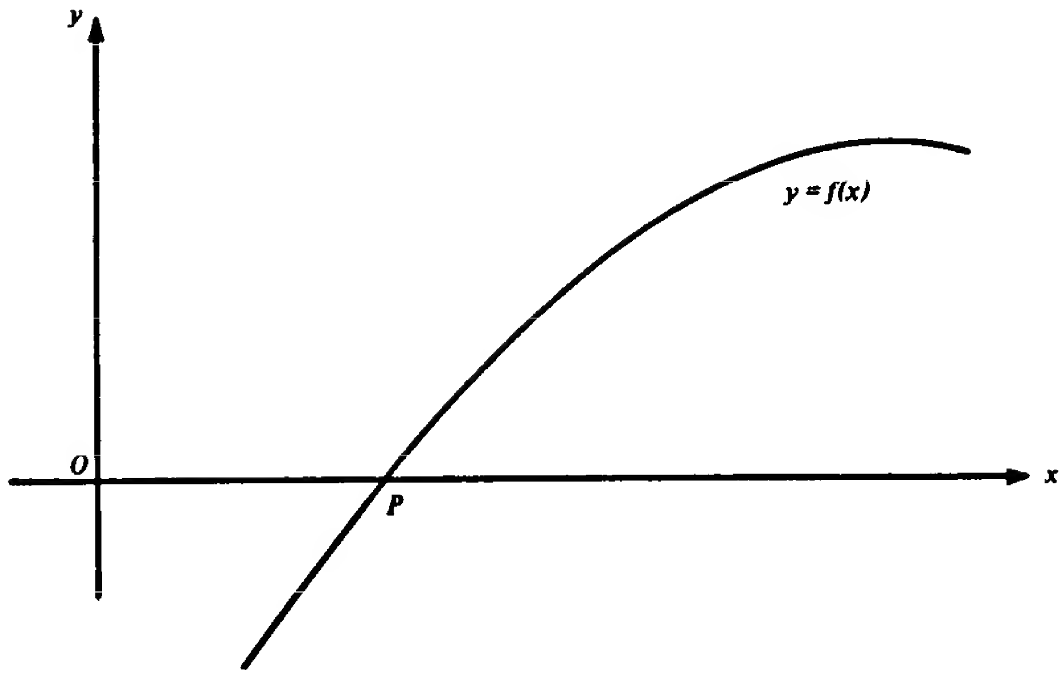

Figure 1.29

$f(x_0) \neq 0$. Thus it turns out that x_0 is merely an approximation to the root. What to do next?

Newton tells us to do something both simple and effective. At P_0, where co-ordinates are $x_0, f(x_0)$, draw the tangent to the curve. Suppose this tangent to meet the x-axis at a point A_1 with abscissa x_1; then x_1 is seen to be a better approximation. See Fig. 1.30.

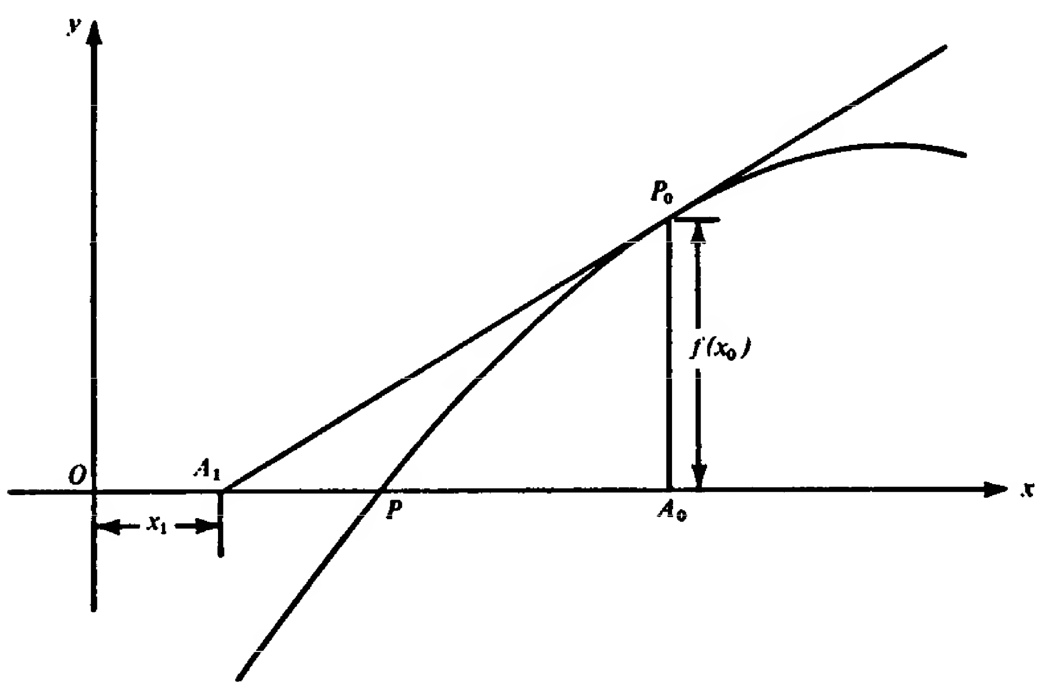

Figure 1.30

Next at P_1, the point on the curve with the abscissa x_1 of A_1, draw a tangent to the curve to cut the x-axis in a point A_2 with abscissa x_2. Repeat the procedure. At P_2, the point on the curve with the abscissa x_2 of A_2, draw a tangent to the curve to cut the x-axis in a point A_3 with abscissa x_3. It is visibly apparent that the resultant sequence of points A_0, A_1, A_2, A_3 are successively nearer to P, so that x_0, x_1, x_2, x_3 are progressively better approximations to the required root. See Fig. 1.31. But though it is evident that in principle the procedure may be repeated until any required degree of accuracy is obtained, in practice there is a limit to the number of repetitions possible with pencil and graph paper. In our diagram the thickness of our pencil line for the tangent at P_3 defeats further accuracy. The role of geometry is to illustrate the method; for its unlimited application we need a formula.

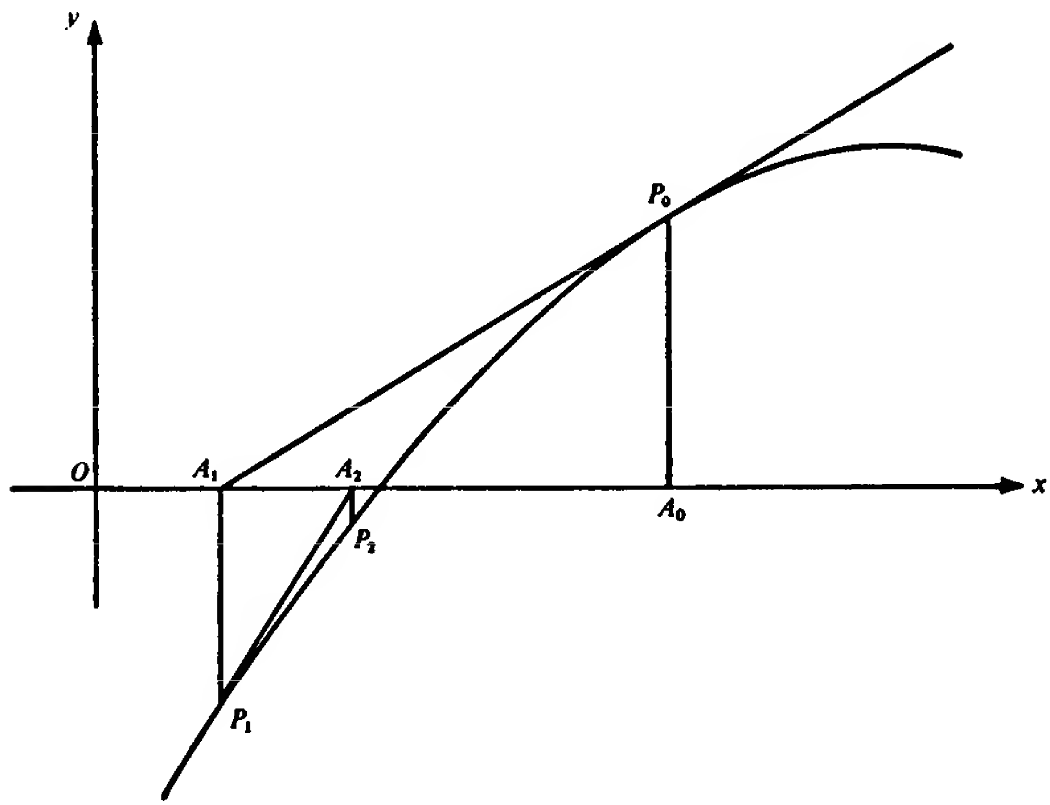

Figure 1.31

1.4.2 Newton's Formula

In $\triangle A_0A_1P_0$, see Fig. 1.32,

$$(1) \qquad \tan\tau = \frac{\text{vertical rise}}{\text{horizontal advance}} = \frac{A_0P_0}{A_1A_0} = \frac{f(x_0)}{x_0 - x_1}.$$

And here the differential calculus comes in useful, for tan τ is also the slope at P_0 of the tangent to the curve $y = f(x)$. We recall that the slope of this curve at the point $P_0(x_0, f(x_0))$ is the value of its differential

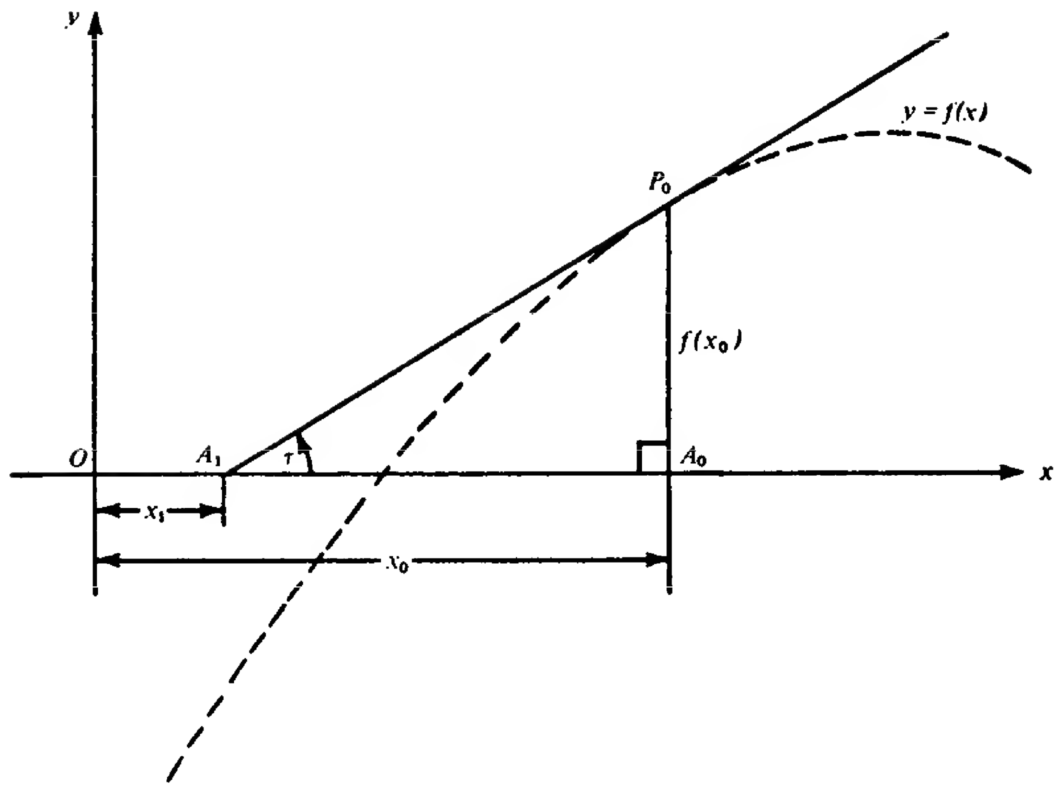

Figure 1.32

coefficient dy/dx when $x = x_0$, or to use an alternative notation, the value of $f'(x)$ when $x = x_0$. The latter is expressed with convenient notational compactness by $f'(x_0)$. So, in short,

$$\tan \tau = f'(x_0). \tag{2}$$

From (1) and (2)

$$\frac{f(x_0)}{x_0 - x_1} = f'(x_0)$$

giving

$$\frac{f(x_0)}{f'(x_0)} = x_0 - x_1,$$

so that

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Given x_0 this formula enables us to compute x_1. Yet the efficacy of Newton's method is its generality; x_2 will be computed from x_1 and x_3

from x_2 in exactly the same way as x_1 from x_0. That is, we have similarly,

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)},$$

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)},$$

and Newton's famous formula, in its full generality, is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Although there are excellent intuitive grounds for supposing Newton's formula to work, for supposing that successive approximations are better approximations, it is nevertheless prudent to test it. To be chary of the untried, hesitant to accept the unproved, is the very first requirement of a scientific attitude. So, let us try it out.

1.4.3 $\sqrt{a}$

Suppose that we wish to find $\sqrt{a}$, the positive root of

$$f(x) = x^2 - a = 0.$$

Here,

$$f'(x) = 2x,$$

so that applying Newton's formula, the right-hand side of our required equation is

$$x - \frac{x^2 - a}{2x}.$$

Remembering to put in the necessary subscripts, the complete equation is

$$x_{n+1} = x_n - \frac{x_n^2 - a}{2x_n}.$$

Hence,

$$x_{n+1} = x_n - \frac{x_n}{2} + \frac{a}{2x_n} = \frac{x_n}{2} + \frac{a}{2x_n}$$

$$= \frac{x_n + a/x_n}{2}.$$

But

$$x_n \cdot \frac{a}{x_n} = a,$$

so that if x_n is an overestimate for $\sqrt{a}$ then a/x_n is an underestimate, and conversely. These considerations are consistent with x_{n+1}, i.e. the mean of x_n and a/x_n, being a better approximation to a than x_n, but do not prove it. So let's get down to brass tacks. Suppose we wish to find $\sqrt{2}$, and that our crude initial approximation (x_n where $n = 0$) is $x_0 = 2$. Then $a/x_n = a/x_0 = 2/2 = 1$, so that

$$x_1 = \frac{2+1}{2} = \frac{3}{2}.$$

Thus

$$\frac{a}{x_1} = \frac{2}{3/2} = \frac{4}{3},$$

and

$$x_2 = \frac{3/2 + 4/3}{2} = \frac{17}{12}.$$

Hence

$$\frac{a}{x_2} = \frac{2}{17/12} = \frac{24}{17},$$

and

$$x_3 = \frac{17/12 + 24/17}{2} = \frac{577}{408}.$$

Our successive approximations x_0, x_1, x_2, x_3 are

$$2, \quad \frac{3}{2}, \quad \frac{17}{12}, \quad \frac{577}{408},$$

whose squares are

$$2+2, \quad 2+\frac{1}{2^2}, \quad 2+\frac{1}{12^2}, \quad 2+\frac{1}{408^2}.$$

Letting the facts speak for themselves is a strong argument.

1.4.4 $\sqrt[3]{a}$

To test Newton's formula further, suppose that we wish to find the cube root of a, the (real) zero of

$$f(x) = x^3 - a.$$

Here

$$f'(x) = 3x^2$$

so that

$$x_{n+1} = x_n - \frac{x_n^3 - a}{3x^2}.$$

Hence

$$\begin{aligned} x_{n+1} &= x_n - \frac{x_n}{3} + \frac{a}{3x_n^2} \\ &= \frac{2x_n}{3} + \frac{a}{3x_n^2} \\ &= \frac{2x_n + a/x_n^2}{3}. \end{aligned}$$

At this stage it is illuminating to take a numerical example, Suppose that we wish to compute $\sqrt[3]{26}$, i.e., to find the real zero of

$$f(x) = x^3 - 26.$$

Evidently 3 is a better approximation than 1, 2, 4, or 5, so let us start with $x_0 = 3$. Consequently $a/x_0^2 = 26/9$. Since

$$3 \cdot 3 \cdot 3 = 27 > 26,$$

3 is greater than $\sqrt[3]{26}$, yet note that

$$\frac{26}{9} \cdot (3 \cdot 3) = 26 = \sqrt[3]{26} \cdot (\sqrt[3]{26} \cdot \sqrt[3]{26}).$$

Consequently $\frac{26}{9}$ is less than $\sqrt[3]{26}$. But $\frac{26}{9} = 2\frac{8}{9}$, so that $\sqrt[3]{26}$ is already pinched between $2\frac{8}{9}$ and 3.

Alternatively, if we take $x_0 = 5$ (say), then $a/x_0^2 = \frac{26}{25}$, which leaves $\sqrt[3]{26}$ between $1\frac{1}{25}$ and 5, with plenty of elbow room, so to speak. We appreciate the convenience of a close initial approximation.

Returning to our computation with $x_0 = 3$, we find that

$$x_1 = \frac{2x_0 + a/x_0^2}{3} = \frac{2 \times 3 + 26/9}{3} = \frac{80}{27}$$

so that

$$\frac{a}{x_1^2} = \frac{26}{(80/27)^2}.$$

But

$$\frac{26}{(80/27)^2} \cdot \left(\frac{80}{27} \cdot \frac{80}{27} \right) = 26 = \sqrt[3]{26} \cdot \left(\sqrt[3]{26} \cdot \sqrt[3]{26} \right),$$

so that if $80/27$ is too large an estimate for $\sqrt[3]{26}$ then $26/(80/27)^2$ is too small, and conversely. But

$$x_1 = \frac{80}{27} = 2.962\ldots,$$

$$\frac{a}{x_1^2} = 26/\,(80/27)^2 = 2.961\ldots,$$

so that $\sqrt[3]{26}$ is already pinched between 2.962 and 2.961, and is therefore 2.96 correct to two places of decimals.

Continuing the computation, we have

$$x_2 = \frac{2x_1 + a/x_1^2}{3} = \frac{2 \times 80/27 + 26/\,(80/27)^2}{3}.$$

It is left to the reader to show that $\sqrt[3]{26}$ lies between x_2 and a/x_2^2, and to calculate it from their numerical values to as many significant figures as is permissible.

1.4.5 $\sqrt[5]{a}$

Although there is an algebraic formula for the solution of the general quadratic, and very complicated formulae for the cubic and biquadratic equations, it is impossible to obtain such a formula for the general equation of degree 5 or higher. In real life, to solve actual problems, we are obliged to proceed by approximations, by arithmetic procedures that give successively better approximations.

To conclude our testing of Newton's formula let us find $\sqrt[5]{a}$, i.e., solve

$$f(x) = x^5 - a = 0.$$

Here

$$f'(x) = 5x^4,$$

so that

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^5 - a}{5x_n^4}.$$

Hence

$$x_{n+1} = x_n - \frac{x_n}{5} + \frac{a}{5x_n^4}$$

$$= \frac{4x_n}{5} + \frac{a}{5x_n^4}$$

$$= \frac{4x_n + a/x_n^4}{5}.$$

Everybody makes mistakes; nobody is infallible. Whereas we cannot avoid making mistakes, we can make checks to avoid leaving them undetected and therefore uncorrected. Do our formulae for x_{n+1} for $\sqrt[2]{a}$, $\sqrt[3]{a}$, $\sqrt[5]{a}$ exhibit some sort of pattern? Yes, the pattern that for $\sqrt[q]{a}$

$$x_{n+1} = \frac{(q-1)x_n + a/x_n^{q-1}}{q}.$$

Doesn't this uniformity increase our confidence in our working?

Proceeding as on previous occasions, we have

$$\frac{a}{x_n^4} \cdot (x_n \cdot x_n \cdot x_n \cdot x_n) = a = \sqrt[5]{a} \cdot (\sqrt[5]{a} \cdot \sqrt[5]{a} \cdot \sqrt[5]{a} \cdot \sqrt[5]{a}).$$

From consideration of this it is evident that if x_n is greater than $\sqrt[5]{a}$ then a/x_n^4 is less than $\sqrt[5]{a}$, and conversely. Consequently $\sqrt[5]{a}$ must lie between x_n and x/x_n^4. And similarly if $x_{n+1} \neq \sqrt[5]{a}$ then the root must also lie between x_{n+1} and a/x_{n+1}^4. Now suppose that $x_n > a/x_n^4$; then

$$4x_n + x_n > 4x_n + a/x_n^4,$$

so that

$$\frac{5x_n}{5} > \frac{4x_n + a/x_n^4}{5};$$

i.e.,

(1) $$x_n > x_{n+1}.$$

But also, since

$$\frac{a}{x_n^4} < x_n,$$

$$4\frac{a}{x_n^4} + \frac{a}{x_n^4} < 4x_n + \frac{a}{x_n^4},$$

and consequently

(2) $$\frac{a}{x_n^4} < x_{n+1}.$$

From (1), (2) it follows that x_{n+1} lies between x_n and a/x_n^4. Also by (1)

$$\frac{1}{x_{n+1}} > \frac{1}{x_n},$$

$$\frac{1}{x_{n+1}^4} > \frac{1}{x_n^4},$$

(3) $$\frac{a}{x_{n+1}^4} > \frac{a}{x_n^4}.$$

What can we conclude from these considerations? If x_{n+1} is an upper approximation the situation is as illustrated by Fig. 1.33. Since $\sqrt[5]{a}$ lies within the inner interval, x_{n+1} is necessarily a closer upper approximation than x_n and a/x_{n+1}^4 a closer lower approximation than a/x_n^4. We thus squeeze out $\sqrt[5]{a}$ with increasing accuracy.

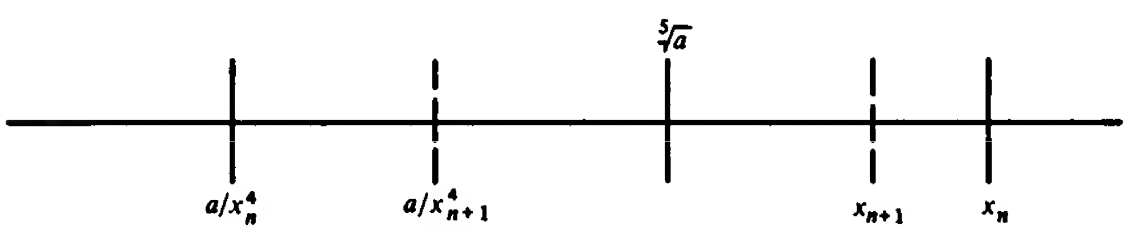

Figure 1.33

CHAPTER TWO

From the History of Statics

Mechanics is the study of the action of forces on bodies. That part in which the bodies are at rest and, consequently, the forces are in equilibrium, is called *statics* in contrast to the other part, *dynamics*, in which the forces are not in equilibrium and, consequently, the bodies not at rest. Here we shall be concerned with the simpler and first-developed branch, statics, which is conveniently introduced by consideration of the contributions of Stevinus and Archimedes. Although the first real achievements are due to Archimedes and preceded Stevinus' by many centuries, I prefer to discuss the latter first.

SECTION 1. STEVINUS AND ARCHIMEDES

Stevinus, a Dutchman, lived in the 16th Century, contemporary with Descartes, a century or so before Newton, Leibniz, and the invention of the differential calculus. He was a brilliant applied mathematician who was fascinated by the usefulness of mathematics: for Stevinus, mathematics to be good had to be good for something. He was one of the first to use decimal fractions and showed their usefulness for everyday affairs, he invented the first horseless carriage, and he constructed dykes, which still serve Holland to this day. His achievements are commemorated by his statue in his native city, Brügge. If you ever go there, look him up. Meanwhile we shall consider his derivation of the Law of the Inclined Plane.

2.1.1 Inclined Plane

Even crude, casual, unavoidable everyday experience presents the curious with questions. Indeed, the simpler the experience the more

difficult to avoid meeting pertinent questions head-on. No matter whether or not it interests us, we all know that it is harder to push an object up a steep incline than up a less steep: the steeper the incline the harder we need push. An incline formed by a pair of planks enables us to slide into our station wagon a trunk too heavy to lift, and for the same good reason the brewer loads his dray by rolling the casks of beer up a ramp. Brains decrease the need for brawn: this simple machine has the merit that the incline takes part of the weight. The curious naturally raise the question: Since pushing up is less strenuous than lifting, what precisely is the saving in effort? It all depends. Yes, but on what? Stevinus was curious.

After pondering these matters, Stevinus conceived the question in a new context. "How does the pull (or push) to move a heavy body up an incline compare with the force necessary to lift it directly?" was asked in relation to the situation here illustrated by Fig. 2.1. Since the tension w in the string counterbalances the force acting down the inclined plane, the ratio pull to direct lift is $w : W$. But a vertical plane is a special case of an inclined plane, so that the underlying general situation is the one illustrated in Fig. 2.2, and the pertinent question, given equilibrium, is: What is the ratio of w to W?

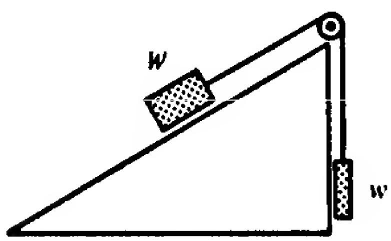

Figure 2.1

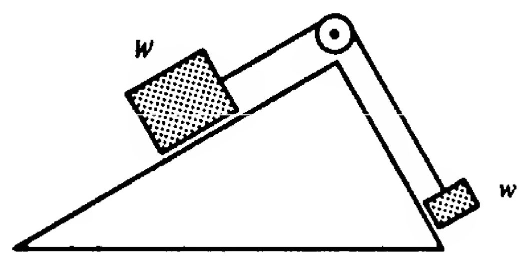

Figure 2.2

Our crude, uncritical, everyday experience suffices to begin an answer. We know that the steeper the slope the greater the pull. When the angle of inclination is zero, no horizontal force is needed to maintain W in equilibrium; when the angle of inclination is 90°, a vertical force W is necessary. Consider the cases of increasing inclinations pictured in Fig. 2.3. Surely for intermediate angles intermediate forces will be required. Were W, w equal, the body on the steeper incline (on the right in Fig. 2.2) would exert the greater pull and consequently slide down. We must conclude that for equilibrium, $w < W$. Yes, but how much less?

We postpone this question to consider several matters of importance raised by the foregoing. First, that varying the data is an important ingredient of puzzling something out. By this method we have come to focus our attention on the fact that the force to maintain a body in

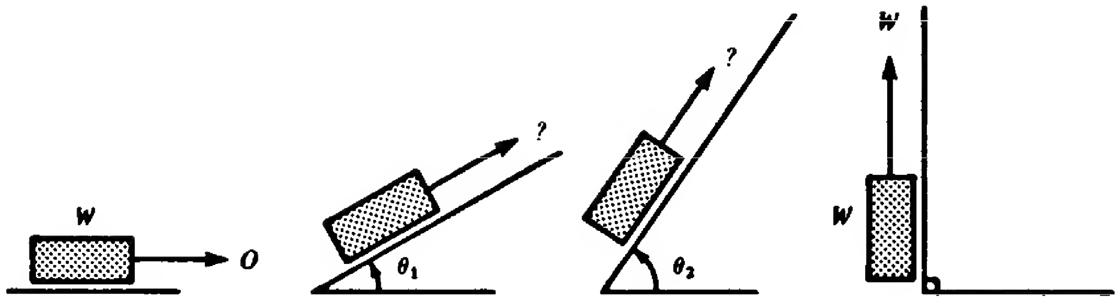

Figure 2.3

equilibrium on an inclined plane is dependent upon the steepness of the incline.

Second, in retrospect many tacit assumptions come to light. We all know that when using a ramp to load a station wagon with a heavy trunk, if we stop to get our breath, the trunk does not necessarily slide back down the ramp. Friction can be sufficient to prevent its sliding down when we cease pushing it up. We said nothing about friction. Next, consider the situation illustrated in Fig. 2.1. That the incline might sag ever so slightly under the weight W was neglected. That friction at the axle of the pulley wheel would resist turning and consequently movement of the string; that the string has weight and therefore the length of it on each side of the pulley wheel is relevant; that the string may not be completely homogeneous, but vary in density; that it is not completely flexible, but offers resistance to bending at the pulley wheel; that the portion between W and the pulley will not be absolutely parallel to the incline, but dependent on its density and flexibility, will sag a little; have been neglected.

Nature is infinitely complex; to render an investigation possible, its complexity must be reduced to manageable proportions. The friction of our station wagon loading ramp can be diminished by making its surface smoother and smoother and by using better and still better quality grease. Finally, with friction reduced to an almost negligible amount, the actuality closely approximates the ideal frictionless state. Similarly, by using thinner, more flexible, more uniform string, more rigid and smoother planes, and better quality pulleys, we minimize the influences of minor circumstances of Fig. 2.1. The closer the actual approaches the idealized state, the more precisely we can test the theory consequent upon our idealization.

Third, there is the point so often made by philosophers in their lectures on the nature of science, that physics is an observational science. It is. But let not this mislead you into supposing that Stevinus first set about solving his problem by precise measurement and critical observation. No. Before

he could intelligently make use of measurements he had to decide what measurements could intelligently be made use of. To the contrary he solved his problem by precise thinking about crude fact. His real problem was conceptual; he had to decide what circumstances were relevant, what irrelevant, and of the relevant, what was of major importance and what could reasonably be neglected. It is precisely this controlled use of the imagination, this conceiving of an idealized situation by abstraction from experience, that is the key to discovery. Until Stevinus had a theory he had no theory to test; his need for precise measurement was subsequent to this theorizing.

We return to the problem of the inclined plane itself. What, with the idealized circumstances of Fig. 2.2, is the ratio of w to W for equilibrium? Deeply pondering this problem, Stevinus appreciated that with no friction equilibrium is independent of the shape of the bodies W, w. Whether these be box-shaped or barrel-shaped is beside the point. Even so, it takes more than an ordinary exercise of the imagination to suppose boxes or barrels to be replaced by rope or chain.

Consider Fig. 2.4. ABC is supposed to be an old-fashioned chain. By "old-fashioned" we mean not today's lean chain with elongated links, but yesterday's fat chain with closely interlocking links that adorned grandfather's watch and waistcoat. This, idealized, is a perfectly flexible metal rope of uniform density. Thus the weights W, w of AB, BC are considered proportional to their lengths, so that the ratio of W to w is that of AB to BC. So our problem, now, is to find the latter ratio. Is there any real prospect of doing so? We seem to have taken a step in the wrong direction.

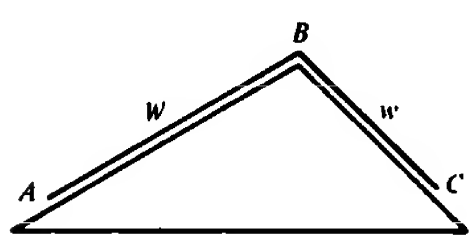

Figure 2.4

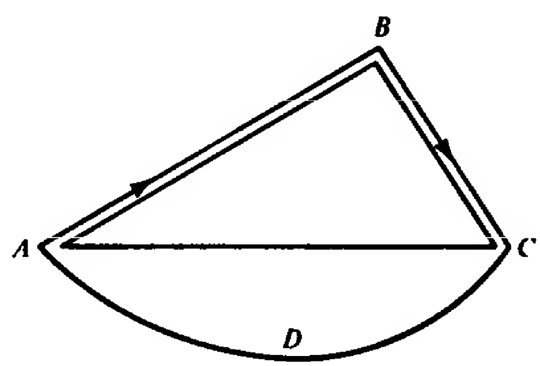

Figure 2.5

The measure of a giant is his stride: Stevinus wore seven league boots. He imagined, what very few of us would imagine, a *closed* chain. Consider Fig. 2.5. Either the flexible homogeneous closed chain hung over the triangular prism is in motion or it is not. Suppose that it is in motion

in, say, the direction ABC. Consider a particle of chain, say that at C. Since it is moving downwards there must be a downward force acting on it. When it has moved, its place at C will be taken by an identical particle. What now? The whole chain still occupies the position it had previously; although each particle has moved a little, each has been replaced by an identical particle; the overall situation remains unchanged. We are forced to concede that if originally there had been a downward force acting on the chain at C, then there still is. Consequently, if the chain is in motion originally, then it is in motion forever. But surely perpetual motion, a free inexhaustible supply of energy, is a philosopher's pipe dream. The Dutch know that from nothing comes nothing; Stevinus was a Dutchman. We conclude that the chain is in equilibrium. And since the whole chain is in equilibrium, the lower portion ADC is. Moreover, the chain being completely flexible, there is no resistance to bending at either A or C, so that it hangs symmetrically below AC. Consequently the downward pull on the particles at A and C are equal; consequently when the lower portion ADC is removed, the upper portion ABC will persist in equilibrum. This situation is illustrated by Fig. 2.6.

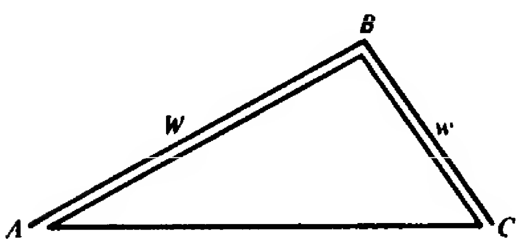

Figure 2.6

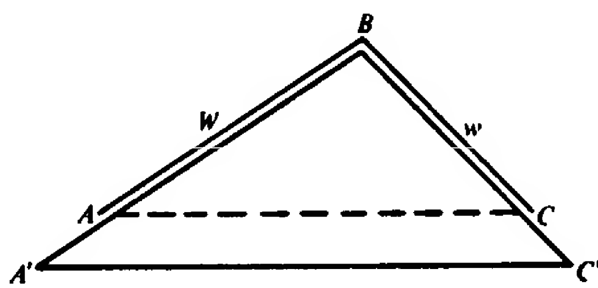

Figure 2.7

But obviously the equilibrium of the chain ABC will be undisturbed if the triangular prism over which it rests is extended. See Fig. 2.7. Whether or not AC is a base edge of the prism does not matter at all. What matters is that AC is horizontal. Suppose, referring to Fig. 2.5, that AC is not horizontal. Since AC is no longer horizontal, the lower portion of the chain does not hang symmetrically below AC, so that the downward forces acting on the particles at A and C are unequal. Consequently when the lower portion is removed, the upper portion ABC is no longer in equilibrium. ABC is in equilibrium if and only if AC is horizontal.

In short, Fig. 2.7 provides the answer to the question of Fig. 2.4. Since, in Fig. 2.7, AC is parallel to $A'C'$, the sides of triangle $A'BC'$ are divided porportionally,

$$\frac{BC}{AB} = \frac{BC'}{A'B}.$$

And since the chain is of uniform density,

$$\frac{BC}{AB} = \frac{w}{W}.$$

Therefore,

$$\frac{w}{W} = \frac{BC'}{A'B}. \tag{1}$$

The ratio of the weights is that of the lengths of the inclines on which the weights rest.

Yet this conclusion holds no matter what the (arbitrary) inclinations of $A'B$ and BC' to the horizontal. Consequently equilibrium will still be maintained even if BC' is vertical, provided only that $A'C'$ remains horizontal. This situation is illustrated by Fig. 2.8. Here,

$$\frac{BC'}{A'B} = \sin \alpha.$$

So that by (1)

$$\frac{w}{W} = \sin \alpha \tag{2}$$

giving

$$w = W \cdot \sin \alpha. \tag{3}$$

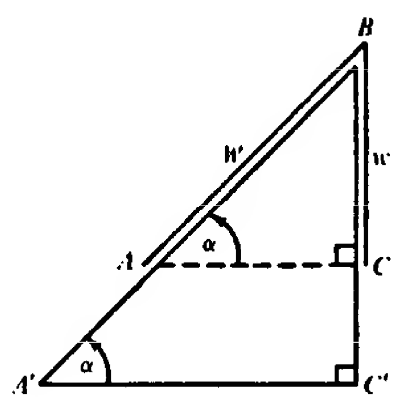

Figure 2.8

It remains merely to remark that the counterbalancing tensions acting at B, and consequently the equilibrium of the system, would be unaltered if the homogeneous chain were replaced by a weightless string with a weight W at its left and a weight w at its right. We conclude that equation (2) is the answer to the problem illustrated by Fig. 2.1 and to the original question "How does the pull (or push) to move a heavy body up an incline compare with the force necessary to lift it directly?"

It is prudent to check conclusions. By (3), when $\alpha = 0°$, $\sin \alpha = 0$, so that $w = 0$, and when $\alpha = 90°$, $\sin \alpha = 1$, so that $w = W$. Stevinus' formula is correct for horizontal and vertical planes. We have reached the stage, subsequent to theorizing, where precise experimental measurement is appropriate—to test the theoretical results for intermediate cases. At this point Stevinus had a theory to test, and tested it. The theory satisfied the examiner.

His solution, obvious to hindsight, requires the foresight of genius. We cannot force ourselves to get such bright ideas.

The substance of this account of the inclined plane as that of the following section on the lever was taken from Ernst Mach's *Principles of Mechanics*, of which a good English translation (1893) of the original German (1883) is available. Mach, besides being an able physicist whose experimental work on sound is commemorated by the Mach unit, was the outstanding philosopher of science of his day. To write this treatise he had first to read Archimedes in the original Greek, Galileo in Italian, Stevinus in Dutch and others in Latin. Modern specialists claim that there are a few points on which he misunderstood the original texts, but when we recall that he was primarily both philosopher and physicist rather than linguist, occasional misinterpretation is only to be expected. Despite minor blemishes, it is a remarkable work by a remarkable man: to me the most fascinating book I have ever read, for I read it at the right time, when young, but not too young. It demands very little mathematics, but lots of common sense. It merits being read several times.

2.1.2 Lever

We are a perverse lot. Although Archimedes (287–212 B.C.) is acknowledged as the greatest of the Greek mathematicians, it is customary not to credit him with what he did do and to credit him with what he did not do. His ingenious methods of computing areas and volumes brought mathematics to the threshold of the integral calculus, yet the textbook gives full credit for the calculus to Newton and Leibniz. He initiated the science of mechanics by discovering the conditions of equilibrium of a lever, yet it often is said that he discovered the lever itself—despite Egyptian pyramid builders using levers thousands of years before he was born.

Here I propose to do no more than introduce the reader to the train of thought underlying Archimedes' discovery of the conditions of equilibrium of a lever. For a complete account of his theory of levers, read Mach.

Although, in considering weights suspended by strings from a beam or lever balanced about a fulcrum, Archimedes never actually says so, context makes it clear that the lever itself is supposed to be rigid and

weightless and the strings weightless and flexible. We find inevitable idealization. His style is mathematical; he begins with explicit statement of his additional, non-contextually implied, assumptions. The first of these, considered so obviously true as to be termed axiomatic, is

Axiom 1. Equal weights at equal distances are in equilibrium.*

It is of course understood that the distances are measured from the fulcrum and that the suspended weights are on opposite sides. Fig. 2.9 illustrates this axiom.

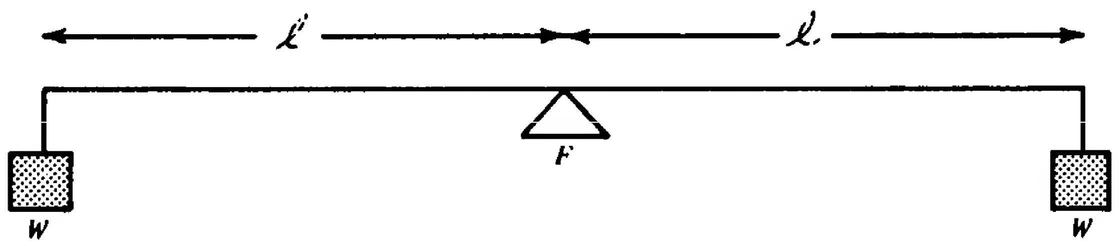

Figure 2.9

This axiom raises two questions. The first: Do we believe it? Is it the right rule for equilibrium of equal weights? But think a moment. There could not be a correct rule or an incorrect rule if there were no rules at all. So there is a second, yet logically prior, question: Are rules possible? It is tempting to retort, "Of course there must be rules." Of course? Must? There is no must about it. We do not know. Yet without rules there could be nothing properly termed science, and with no science to pursue there could be no pursuit of science. We take it as an article of faith that science is possible, that there are rules.

Let us return to the first question: Is Axiom 1 the correct rule for equilibrium of equal weights? Obviously. We all know how to weigh a pound of bacon with a pair of (equi-armed) scales. Archimedes has merely made articulate our common experience. So his rule is "obvious" in the sense that we are familiar with its exemplifications. And we are all familiar with boiling water changing into steam; obviously boiling water makes steam. That it happens is obvious; why it happens is not obvious. That Axiom 1 applies to scales is obvious; why it is applicable is not.

This brings us to the Principle of Sufficient, or if you prefer, Insufficient, Reason. This principle is illustrated by the story of Buridan's ass. Buridan was a scholastic philosopher who is nowadays remembered only because of his ass—even though it is far from certain that the story of Buridan's ass is Buridan's story. But no matter whose ass it was, the poor

*The Works of Archimedes, edited by T. L. Heath (Dover), p. 189.

creature found himself equidistant from two identical bundles of hay. Symmetrically placed between these equally sweet-smelling bales, the poor ass could find as much reason to go first to the one as first to the other and no more reason to go first to the one than first to the other. And so, as a consequence of the Principle of Sufficient, or Insufficient, Reason, it died of hunger.

We turn from Buridan's ass to Archimedes' lever. Lever, strings, and weights being symmetrically situated with respect to the fulcrum, there is as much and as little reason why the right weight should sink as the left should. Suppose that the right-hand weight sinks. But which weight is the right-hand weight? View the lever from the other side and the side previously said to be the right must now be described as the left. Thus a right-hand rule is inconsistent. So, similarly, is a left-hand rule. Such rules depend upon the point of view of the observer, yet the lever does not care whether it is observed or not. The only consistent alternative is Axiom 1.

Archimedes makes a second explicit assumption. It may have been suggested by the following common experience. We all know that it is easier to carry a ladder with help than to carry it alone. Unassisted, you take the whole weight on your shoulders; assisted, you share the weight with other shoulders. Consider carrying a (uniform) ladder with a fellow ladder carrier, one of you at each end. Who takes the greater weight? Change ends. As far as your shoulders can tell you take the same weight as before; you share the weight equally. Thus we are led to argue that in the idealized case where the ladder carriers are twins with shoulders the same height above the ground and so forth, the situation is perfectly symmetrical, so that each pair of shoulders takes exactly half the ladder's weight. Carry the ladder without help and you put your shoulder to its midpoint to balance it.

Let us turn our attention from supporting shoulders to supported weights. We conclude that the equilibrium of a weightless ladder, rod, or beam with a weight W suspended from each end, will be undisturbed by replacing both weights by a single weight $2W$ suspended from the ladder's midpoint. And conversely of course, $2W$ at the midpoint may be replaced by W at each end without destroying the equilibrium. This is (essentially) Archimedes' second assumption. The context understood, we may put it tersely as

$$\text{(A)} \qquad W \text{ at each end} \equiv 2W \text{ in the middle} \qquad \text{(with equilibrium).}$$

This assumption is illustrated by Fig. 2.10.

From Axiom 1 and Assumption (A), or rather from a generalization of the first, Archimedes' Law of the Lever is deducible. I shall give some

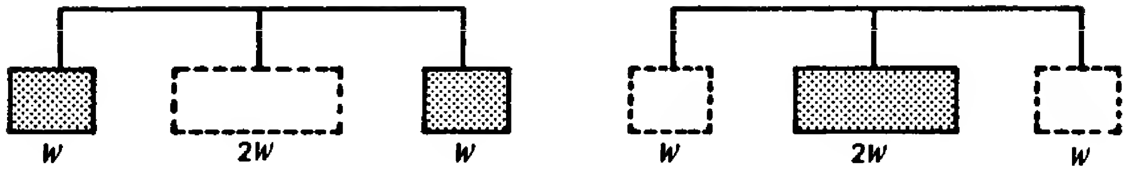

Figure 2.10

insight of the method of proof for the general case by considering specific examples.

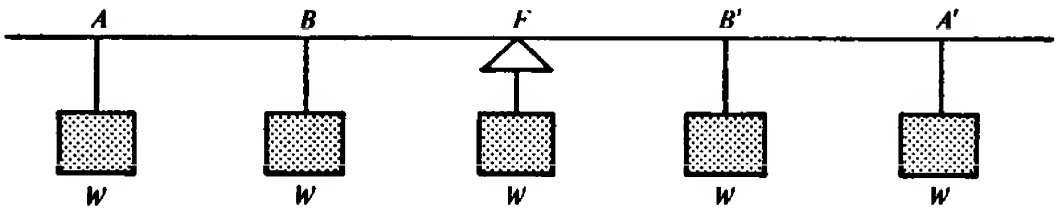

Figure 2.11

First, study Fig. 2.11. The five equal weights are supposed spaced at equal intervals, say, unit distance. The whole system is symmetrically placed with respect to the fulcrum and so, by the principle of insufficient reason, in equilibrium. We have an alternative argument. Since by Axiom 1 the weights at A, A' would in the absence of all other weights ensure equilibrium, while similarly the weights at B, B' would in the absence of all other weights ensure equilibrium, as would that at F, we conclude that the weights at A, A', B, B', and F together ensure equilibrium.

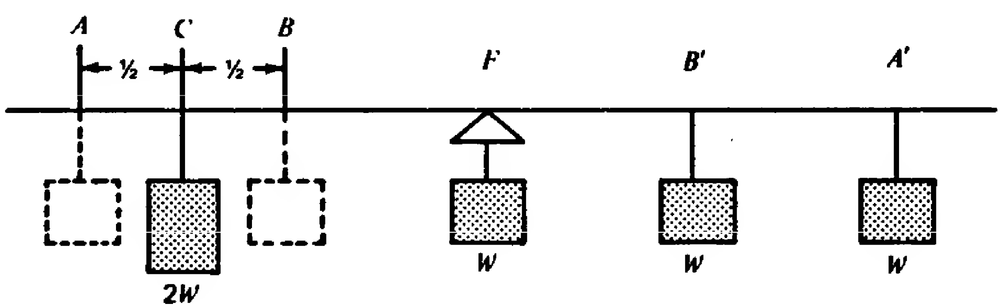

Figure 2.12

Next, study Fig. 2.12. By Assumption (A) the equilibrium of the segment AB of the lever is unchanged when W at A and W at B are replaced by $2W$ at C; consequently the equilibrium of the entire lever is unchanged.

And finally, study Fig. 2.13. Again using Assumption (A), equilibrium of FA' is unchanged when W at F and W at A' are replaced by $2W$

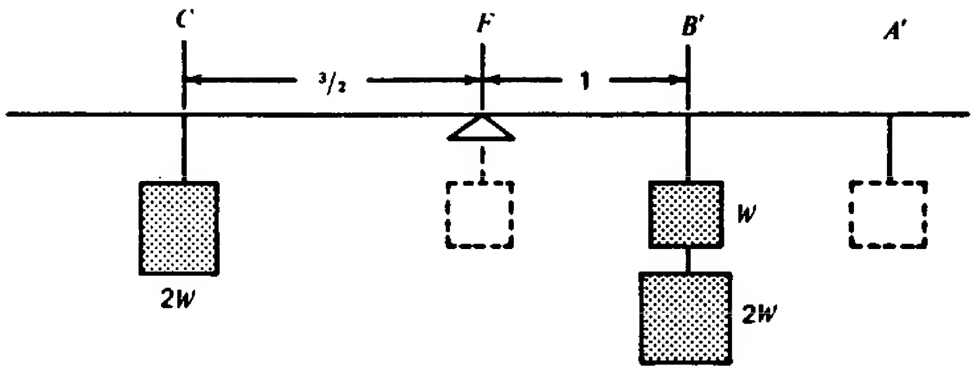

Figure 2.13

at B'; consequently the equilibrium of the entire lever is unchanged. In short, we conclude that a weight of $2W$ suspended 1.5 units from the fulcrum F will balance a weight of $3W$ suspended 1 unit from its other side. But

$$(1) \qquad 2W \cdot 1.5 = 3W \cdot 1;$$

i.e.,

weight · distance from fulcrum = weight · distance from fulcrum.

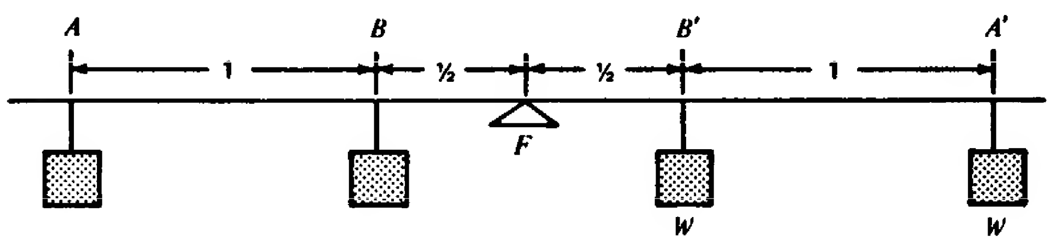

Figure 2.14

Let us consider another special example. See Fig. 2.14. By symmetry, or by using Axiom 1, we conclude that the lever is in equilibrium.

Next, see Fig. 2.15. By Assumption (A) the equilibrium of BA' is unchanged when W at B and W at A' are replaced by $2W$ at B'; consequently the equilibrium of the entire lever is unchanged. Thus W acting at $\frac{3}{2}$ units from the fulcrum balances $3W$ acting at $\frac{1}{2}$ unit from it. But

$$(2) \qquad W \cdot \tfrac{3}{2} = 3W \cdot \tfrac{1}{2};$$

i.e., again

weight · distance = weight · distance.

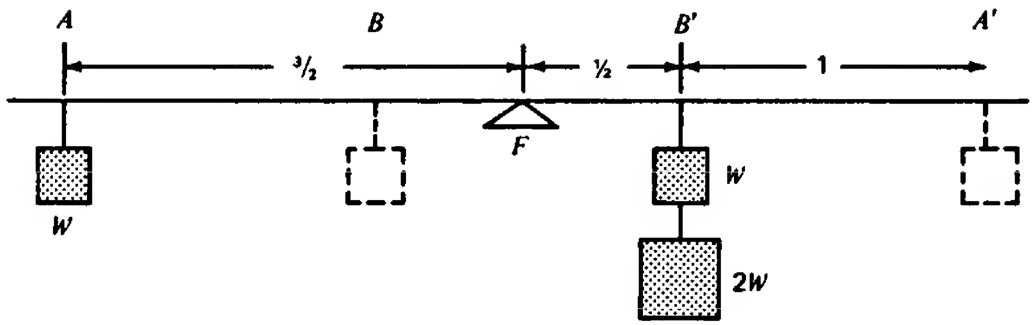

Figure 2.15

Note that multiplying (2) by 2 gives (1) or

$$W \cdot \left(2 \cdot \tfrac{3}{2}\right) = 3W \cdot \left(2 \cdot \tfrac{1}{2}\right),$$

that is,

$$W \cdot 3 = 3W \cdot 1.$$

Illustrate this alternative interpretation and use Archimedes' axiom to show that equilibrium is obtained. We can conjecture on the same general conditions for the equilibrium of levers.

SECTION 2. VECTORS

The notion of a vector arises quite naturally and is basic to physics and indispensable to applied mathematics. That it is clear from the outset that vectors are good for something makes the topic readily teachable at an elementary level. That vectors are becoming part of the high school program is a real step forward.

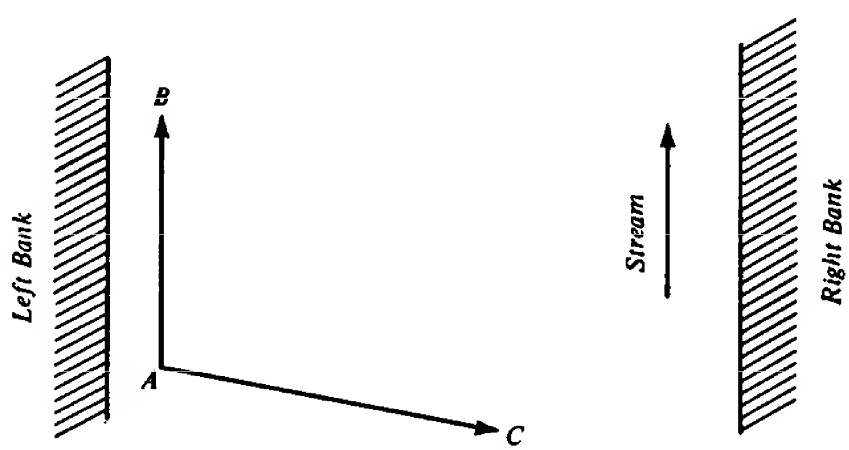

Figure 2.16

We begin with an example. A man is to cross a river from the left bank to the right. Too lazy to row, he uses a motor boat. If his motor fails to start when he casts off, he will drift down river with the tide. Let us suppose him to drift AB in unit time. See Fig. 2.16. If it is high tide so that there is neither a current up nor down river, and his motor is working, he will travel, let us say, AC in unit time. But, if both tide and motor are working, his boat will have velocities due to both. Where will it be at the end of unit time?

The answer comes quite naturally. Consider a special case. A boat at A heading up river at, say, 440 feet per minute (that is 5 m.p.h.) against a down-river current of the same velocity moves neither up nor down river; with both velocities simultaneously it stays put relative to the river bank. At the end of a minute it is in the same position as it would be at the end of two minutes if it moved solely under the influence of the current with no motor for the first minute and under the influence of the motor with no current for the second minute. In the first minute it would move 440 feet down river with the current and in the second minute motor 440 feet back up the (now currentless) river. Thus (at the end of two minutes) it would be in the same position after current and motor acted *successively* (for a minute each) as it would be after both acted *simultaneously* (for a minute). In short, the resultant effect of both forces, current and motor, is that of each acting independently of the other.

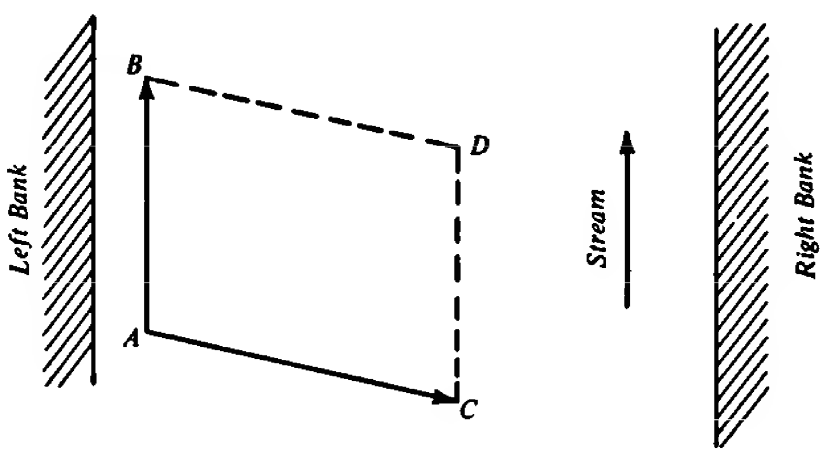

Figure 2.17

Thus, returning to the general case of Fig. 2.16, it is natural to suppose that the boat will at the end of unit time, say, a minute, be at D, where $ABDC$ is a parallelogram. See Fig. 2.17. In one minute the boat acted on by current without motor would drift to B; in the succeeding minute acted on by motor without current it would go as far as (and in the same direction as) if it started from A instead of B, i.e., from B to D (instead of from A to C). So, under the forces due to current and motor

acting successively, at the end of two minutes, it is at D. Alternatively conceived, the boat starting at A, acted upon by motor without current would in one minute reach C, and in the succeeding minute acted upon by current without motor, would drift down river a distance (from C) equal to AB, i.e., it would drift from C to D. Viewed either way the successive effects of current and motor (each acting for one minute) is for the boat to reach D. Is it not natural to conclude that the simultaneous effect (for one minute) is for the boat to reach D? We thus arrive at the Parallelogram Law of Displacements.

In half (or double) the time the boat's displacement down river will be half (or double) AB, say AB', and its "across" river displacement half (or double) AC, say AC', so that the boat's position resultant from both displacements will be D' where $AB'D'C'$ is a parallelogram of sides half (or double) those of parallelogram $ABDC$. See Fig. 2.18.

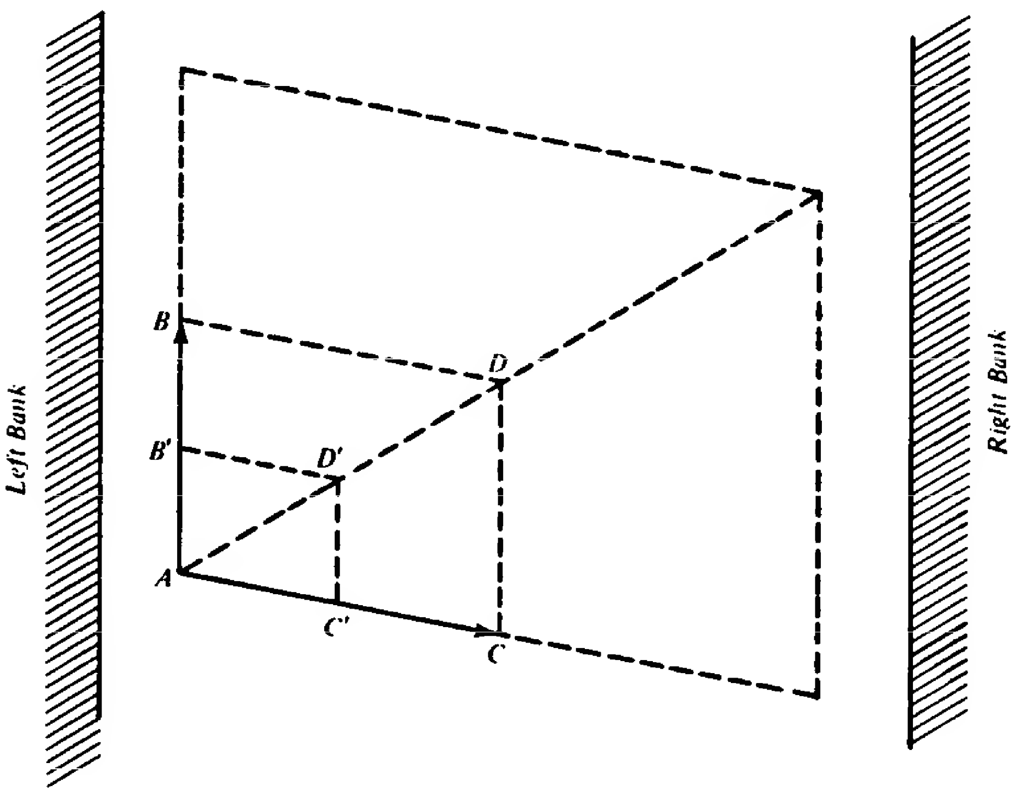

Figure 2.18

More generally, since no matter what the time in question, the ratio of AB' to AB must be the same as that of AC' to AC, the position D' resultant from both the displacements AB' and AC' will be such that parallelograms $AB'D'C'$, $ABDC$, and hence triangles $AC'D'$, ACD are similar. It follows by obvious geometry that D' lies on AD or AD produced. We conclude that the path of the boat is actually along AD. But AB and AC are the distances the boat goes down and "across" river in unit time, one minute, so that these displacements represent its

component velocities in these directions, and AD represents their resultant. We have the Parallelogram Law of Velocities.

Displacements and velocities are remarkable quantities. In addition to having an amount or magnitude they have a direction or sense, so that it is natural to represent them by directed line segments, or, as we say, vectors. The direction of the quantity is indicated by the direction of the line segment, the magnitude of the quantity by the length of the line segment. Precisely because displacements and velocities are both vector quantities, the resultant of a pair of either is represented by the diagonal of the parallelogram through the common point of the sides representing the pair. Many important physical quantities are of the same nature. Boxers know the difference between receiving a straight right and an uppercut; the direction of the blow can be crucial. We must anticipate a Parallelogram Law of Forces.

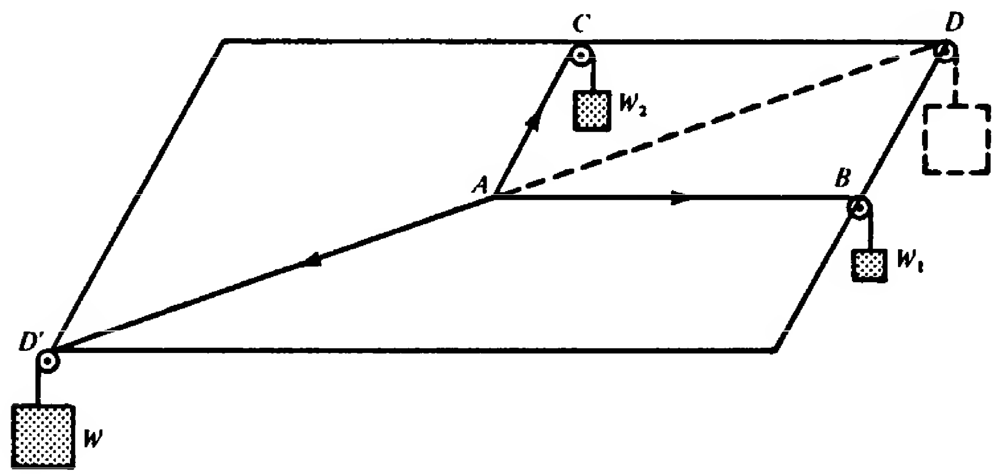

Figure 2.19

Consider the situation illustrated by Fig. 2.19. The particle at A is in equilibrium under a force W_1 along string AB, a force W_2 along string AC, and a force W along string AD'. Since A is in equilibrium under the action of these three forces, it must be in equilibrium under any one of them and the resultant of the other two; in particular, A must be in equilibrium under the action of the force along AD' and the resultant of the forces along AB and AC. But it is clear that A will be in equilibrium only if this resultant is equal in magnitude to the force along AD' and acts in the opposite direction. See Fig. 2.20. Experiment confirms our expectations. It is found that if AB, AC, and AD' are of length W_1, W_2, and W units respectively, then the fourth vertex D of the parallelogram with sides AB, AC is such that AD is W units and D', A, D, are collinear. In short, if AB, AC are vectors representing component forces in magnitude and direction then the diagonal AD of

parallelogram $ABDC$ is a vector representing their resultant in magnitude and direction.

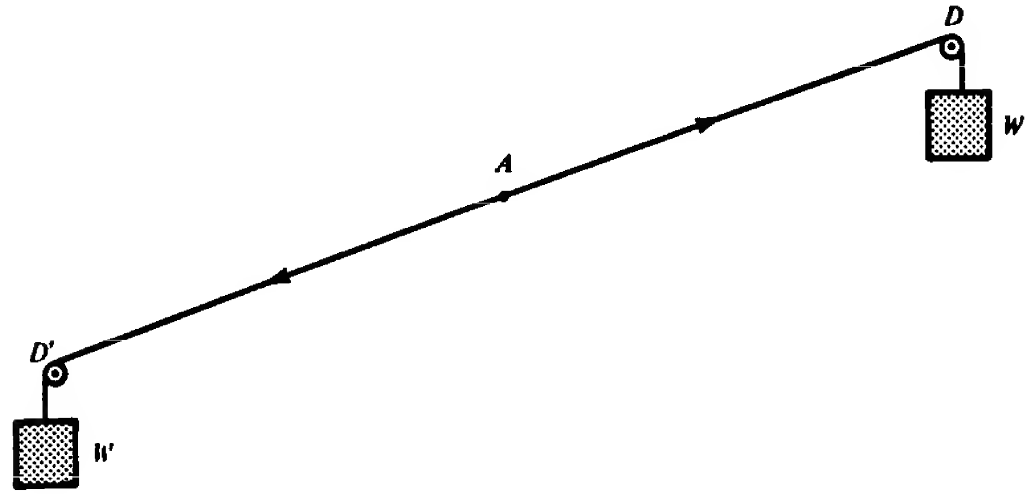

Figure 2.20

Of course an element of idealization is attendant upon this experiment as on any other. In supposing the weights to exert forces W_1, W_2, W on A, we assume the strings to be weightless and perfectly flexible, the little pulley wheels to be frictionless, and so forth. The nearer actual conditions are made to approximate to the ideal, the more exactly is the Parallelogram Law verified.

2.2.1 Inclined Plane

Consider a body of weight W in equilibrium on a rigid frictionless inclined plane of angle α, as illustrated by Fig. 2.21. With the usual idealization, we infer that the tension in the string is w throughout. What is w in terms of W?

The body is in equilibrium under the action of three forces, its own weight W acting vertically downwards, the tension w in the string

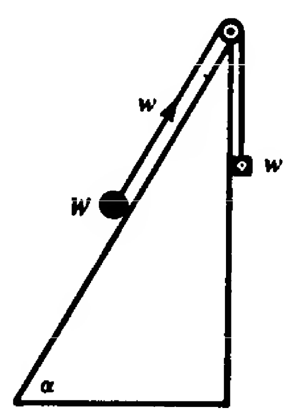

Figure 2.21

acting up the plane, and R the reaction of the plane. Since we suppose there to be no friction between the body and the plane, R cannot have a component force along the plane; R must be normal to the plane. Also, R must be a force of opposite direction (but equal magnitude) to the resultant of the other two forces. But, by the Parallelogram Law of Forces, the direction of the resultant of the other two forces (as well as its magnitude) is represented by the diagonal through A of the parallelogram whose sides with common vertex at A represent W and w. Consequently, the diagonal through A is normal to the inclined plane.

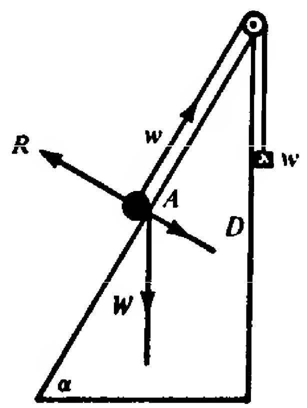

Figure 2.22(a)

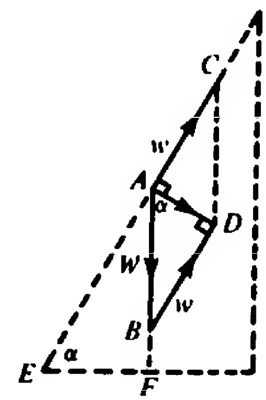

Figure 2.22(b)

See Fig. 2.22(a). From Fig. 2.22(b) we have that $\angle BAD$, $\angle AEF$ are each complementary to $\angle EAB$, so that $\angle BAD = \alpha$, and that since $BD \| AC$, $\angle BDA = 90°$. Hence, since $AB = W$ units, $BD = AC = w$ units. Considering $\triangle ABD$,

$$\frac{w}{W} = \sin \alpha$$

so that

$$w = W \cdot \sin \alpha.$$

Although Stevinus found this result in a most excitingly original way, his underlying principle has the disadvantage that it is far less readily applicable to other problems than the Parallelogram Law.

2.2.2 Pulley

A system of pulleys enables us to lift weights too heavy to lift by unaided muscle power. Suppose, for example, you must remove the engine from your car for a major overhaul. Rather than try to lift it, you

could less strenuously hoist it out by the pulley system illustrated by Fig. 2.23.

As usual, we suppose idealized circumstances; that pulling at C will lift up the engine rather than bring down the roof. With frictionless pulley wheels (the center of A being fixed in position) and weightless rope, a downward force of w at C will give the rope a tension w throughout, so that B when in equilibrium will be in equilibrium under two upward forces of w and a downward force of W. Thus,

$$w + w = W$$

giving,

$$w = \tfrac{1}{2}W.$$

With any increase in w, the engine is hoisted.

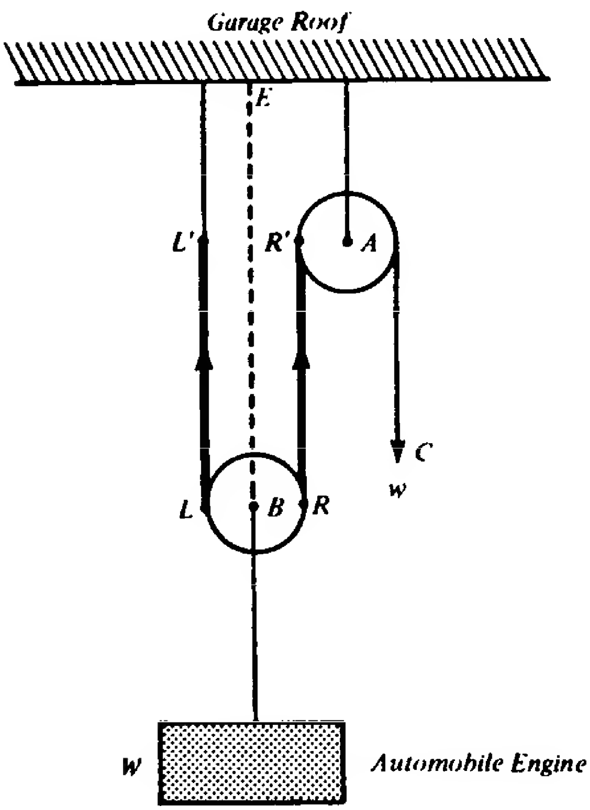

Figure 2.23(a)

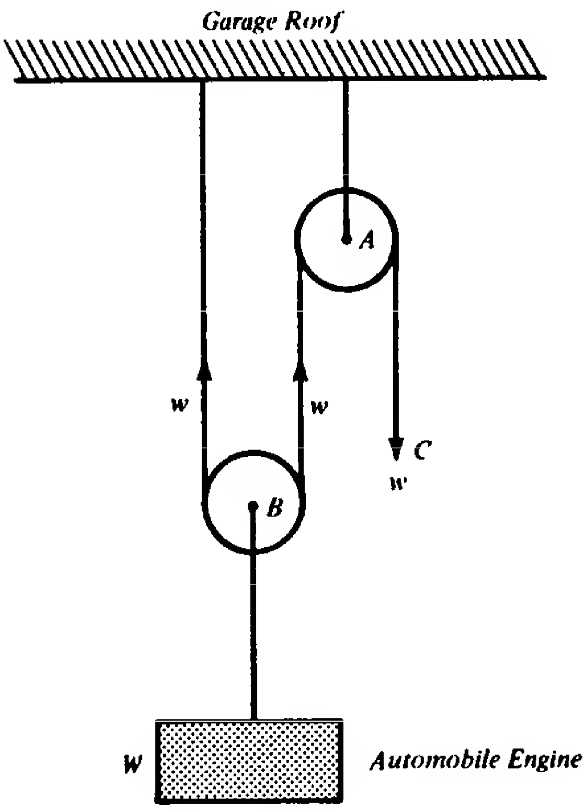

Figure 2.23(b)

Note that this result is a consequence of the Parallelogram Law of Forces if we neglect the dimensions of the pulley with center B. See Fig. 2.23(b), where LL' and RR' (L is for left and R for right) represent in magnitude and direction the upward vertical forces acting on the pulley with center B. When this pulley shrinks to point B, L and R coincide at B, and LL' and RR' become equally inclined at an angle

(say) θ to the vertical BE. We complete the parallelogram $BR'DL'$. See Fig. 2.24.

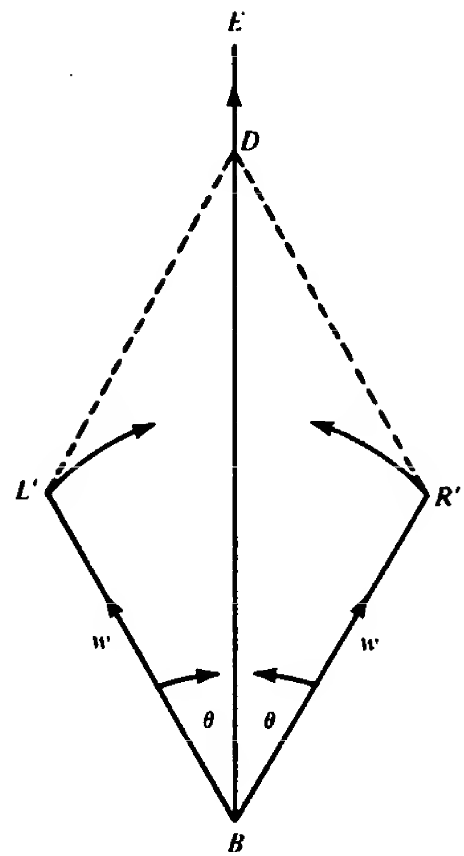

Figure 2.24

When BL', BR' collapse onto BE, L' and R' become coincident, and $R'D$, since it must remain parallel to BL', will lie on BE. But $R'D$ must remain equal in length to BL', so that BD will lie along BE and will be twice the length of BL'. Thus, by the Parallelogram Law the resultant is a force $2w$ acting vertically upwards.

It is left to the reader to show by means of the Parallelogram Law that the resultant of two equal but opposite forces is zero.

2.2.3 Lever

We already have some idea of how Archimedes deduced his Law of the Lever. Let us derive this by applying the Parallelogram Law.

But first a word about rigid bodies. It is evident that a rigid body B will be in equilibrium under the action of two equal, but opposite, forces acting on the same particle of it, say, that at A. See Fig. 2.25. Yet we all know that in a tug of war, different members of our team pull on the rope at different places. Does this matter? Of course not. B will still be in equilibrium if the points of application of the equal and opposite forces

are at, say, A' and A'', instead of both at A. See Fig. 2.26. It does not matter at all where the points of application are, provided that the two equal opposing forces have the same line of action. The transmissibility of the forces at A' and A'' is due to the rigidity of B.

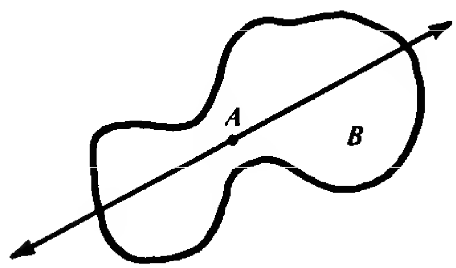

Figure 2.25

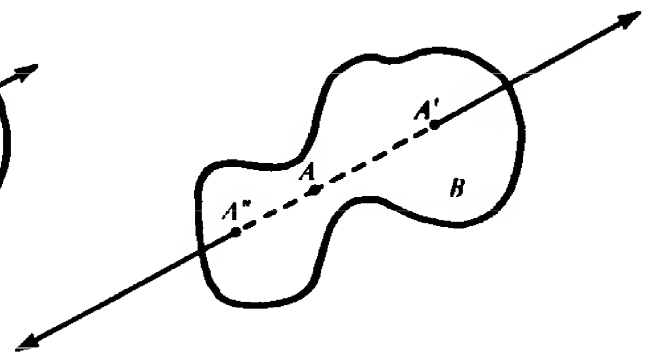

Figure 2.26

The reader will more fully appreciate the importance of this principle of transmissibility when he has seen how it enables us to deduce the conditions of equilibrium of the lever from the Parallelogram Law. To this deduction we now turn.

The general problem may be stated as follows. What are the conditions for the equilibrium of a rigid weightless lever AA' with weights w, w' suspended from A, A'? At what point of AA' is the fulcrum F, and what force must F exert on the lever? Archimedes' Assumption (A), illustrated by Fig. 2.10, suggests part of the answer. In this special symmetrical case where w and w' are equal, an upward force equal to the sum of the weights, acting at a fulcrum at the midpoint of AA', receives equilibrium. Does not this suggest in the general case an upward force of $w + w'$ at some point F in AA'? Yes, but which point? When w and w' are unequal, symmetry is destroyed, F is not the midpoint. Our introduction to Archimedes' treatment of the lever should enable the reader to anticipate the specification of F.

To apply the Parallelogram Law to determine the resultant of w suspended at A and w' at A', we represent these forces by lines $AB, A'B'$, drawn vertically downwards, of w and w' units respectively, thereby representing these forces in both magnitude and direction. Immediately we are confronted by a difficulty. Since $AB, A'B'$ are parallel lines, no matter how far they are extended they cannot intersect; we cannot construct a parallelogram to obtain their resultant. See Fig. 2.27.

Figure 2.27

The difficulty is readily overcome. If the forces w, w' were not parallel, there would be no difficulty. We must substitute equivalent forces (i.e., forces with the same effect) that are not parallel. Can we compound a force with w and a force with w' to give non-parallel resultants with the same effect as w and w' now have? Suppose that at A we introduce two equal but opposite forces; one in the direction $A'A$, the other in the direction AA'. Each of these annuls the effect of the other; equilibrium is undisturbed. But, by the principle of transmissibility, the point of application of one of these may be transferred to A' provided that its magnitude and direction are unaltered. Neither force wins the tug of war. The vector representation of the new situation is given by Fig. 2.28. The vectors $AC, A'C'$ represent the "tug-of-war" forces. Completing the vector parallelograms at A and A', we obtain the vector representations $AD, A'D'$, of forces acting at A and A' respectively. These resultants together have the same effect on the lever as their pairs of components; their pairs of components have the same effect as the original forces. Consequently, if the lever were in equilibrium originally, it still is. We have surmounted our difficulty.

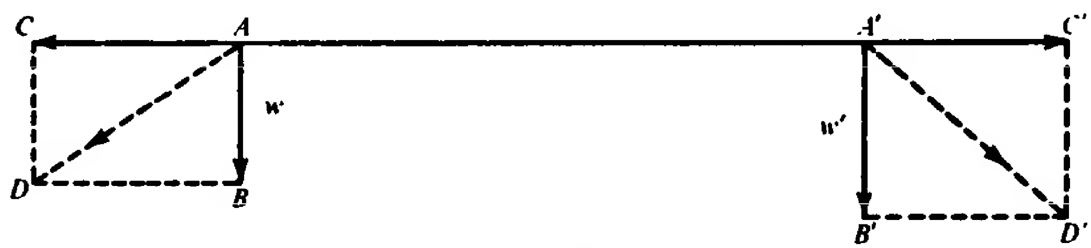

Figure 2.28

Introducing further idealization, let us suppose the lever AA' to become an extended rigid, but weightless, body. Being weightless, no new forces are introduced, so that equilibrium is undisturbed; being rigid, forces acting on it are transmissible, so that their points of application may be moved, collinear with their directions, without change of effect. See Fig. 2.29. Let the lines of action $AD, A'D'$, of the forces at A, A' (produced backwards), intersect at M (we call the point of intersection

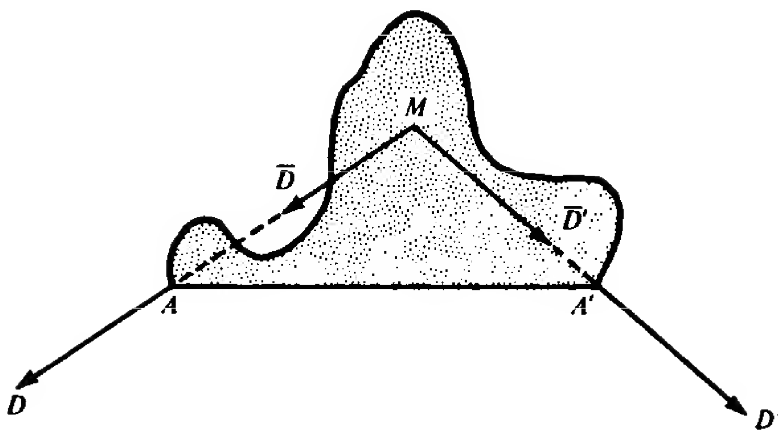

Figure 2.29

M, although M will not in general be above the middle of AA' — M is the meeting point). Then we may, without change of effect, replace the forces whose points of application are A and A' by forces of the same magnitude and direction whose common point of application is M. The vectors of the forces acting at M are the directed lines $M\bar{D}$, $M\bar{D}'$, where of course, $M\bar{D} = AD$ and $M\bar{D}' = A'D'$. See Fig. 2.30.

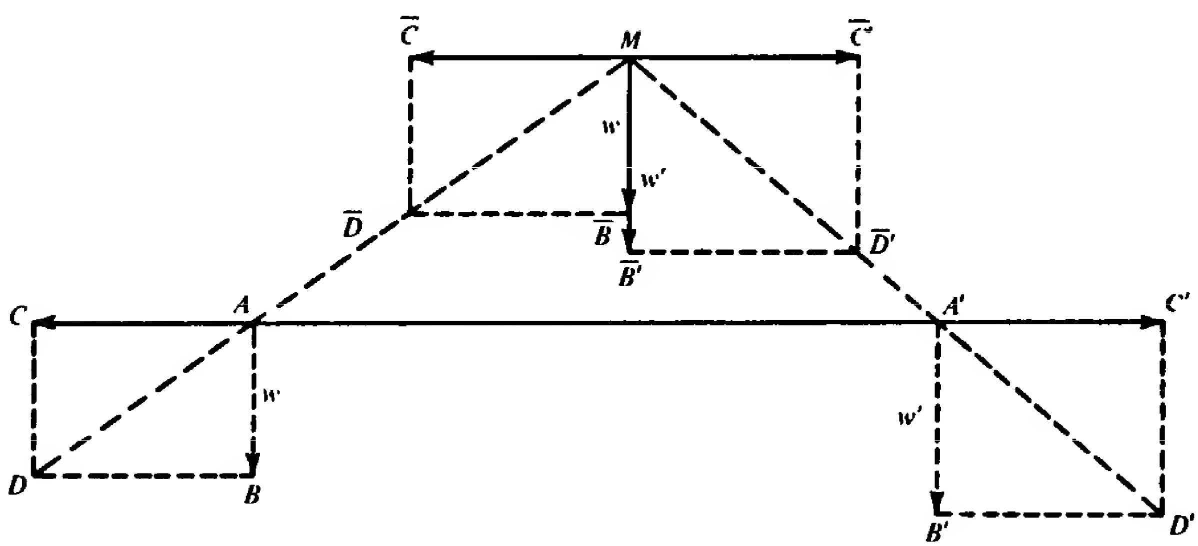

Figure 2.30

But $M\bar{D}$ may be resolved into two component vectors identical with the component vectors of AD (except for point of application) for $M\bar{D}$ has the same magnitude and direction as AD. Similarly, $M\bar{D}'$ has the components of $A'D'$. See Fig. 2.30. But the tug-of-war forces represented by AC, $A'C'$ annulled one another; consequently the pair represented by $M\bar{C}$, $M\bar{C}'$ also annul one another. Thus the resultant force acting at M is the resultant of forces w and w' (equal to those originally at A and A') acting vertically downwards. So, by the Parallelogram Law the resultant force at M is $w + w'$ acting vertically downwards.

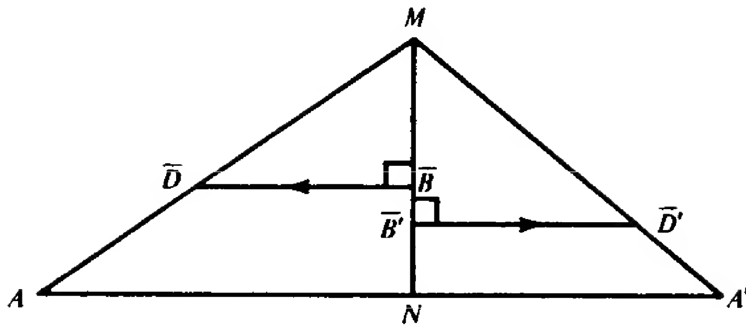

Figure 2.31

The interesting question is: Where does the vertical line of action of this resultant cut AA'? Consider Fig. 2.31. We recall that a line parallel to the base of a triangle divides the sides proportionally. So, considering $\triangle AMN$,

(1) $$\frac{NA}{MN} = \frac{\bar{B}\bar{D}}{M\bar{B}},$$

and considering $\triangle A'MN$,

(2) $$\frac{NA'}{MN} = \frac{\bar{B}'\bar{D}'}{M\bar{B}'}.$$

From (1),

(3) $$NA \cdot M\bar{B} = MN \cdot \bar{B}\bar{D}.$$

From (2),

(4) $$NA' \cdot M\bar{B}' = MN \cdot \bar{B}'\bar{D}'.$$

But

$$\bar{B}\bar{D} = M\bar{C} = AC = A'C' = M\bar{C}' = \bar{B}'\bar{D}'$$

(as is illustrated by Fig. 2.30), so that

$$MN \cdot \bar{B}\bar{D} = MN \cdot \bar{B}'\bar{D}'.$$

Hence from (3) and (4)

(5) $$NA \cdot M\bar{B} = NA' \cdot M\bar{B}'.$$

We have the answer to our question. The vertical line of action of the resultant cuts AA' at a point N such that (5) holds.

But, again by the principle of transmissibility, the replacement of the resultant acting at M by a force of the same magnitude and direction acting at N leaves the effect on the lever (or extended rigid, but weightless body) unchanged: we have shown that the resultant of the original forces w and w' acting vertically downwards at A and A', respectively, is a force $w + w'$ acting vertically downwards at N. But of course the lever would be in equilibrium under two forces of $w + w'$ at N, the one acting vertically downwards and the other vertically upwards. Consequently, if a fulcrum F (sufficient to support $w + w'$) is introduced at N, then the lever is in equilibrium under the original forces. That is, by (5), as $F = N$, the lever will be in equilibrium if

$$FA \cdot M\bar{B} = FA' \cdot M\bar{B}'.$$

Recalling that $M\bar{B} = AB = w$, $M\bar{B}' = AB' = w'$, and putting $FA = a$, $FA' = a'$ (where a stands for *arm*), we have, finally,

$$a \cdot w = a' \cdot w'.$$

We have used the Parallelogram Law of Vectors to derive Archimedes' Law of the Lever.

2.2.4 Archimedes' Application of his Law of the Lever

We have already seen, in rough outline, how Archimedes discovered his Law of the Lever. As you know, he was the greatest of the Greek mathematicians; indeed, he was one of the greatest mathematicians who ever lived. The chief basis of his fame is his discovery of the beginnings of the integral calculus, a discovery brought about by a most ingenious application of his Law of the Lever.

The aims, the most visible aims, that gave rise to the integral calculus, are those of computing areas and volumes which are enclosed, not by straight lines like polygons nor by planes like polyhedra, but by curved lines and curved surfaces. For instance, a problem demanding integral calculus is determination of the volume of a sphere. Not only is it the most natural and most exciting problem about volume; also it is one of the most difficult. Archimedes was the first to solve it. Why is it so difficult? What on earth has this to do with levers?

One question at a time. Why so difficult? Compare the sphere with other volumes, say the cylinder and cone. Whereas the sphere is round in every direction, the latter are, so to speak, merely half round. The lateral surface of a right cylinder can be cut along the straight-line element AB,

peeled off, and flattened out into a rectangle without distorting it. See Fig. 2.32. Similarly the lateral surface of a right cone can be cut along the straightline element OB, peeled off, and flattened out into a circular sector without distortion. See Fig. 2.33. Not so, the surface of a sphere. We have all peeled oranges. That the sphere has a more complicated sort of surface suggests the computation of its volume to be more difficult.

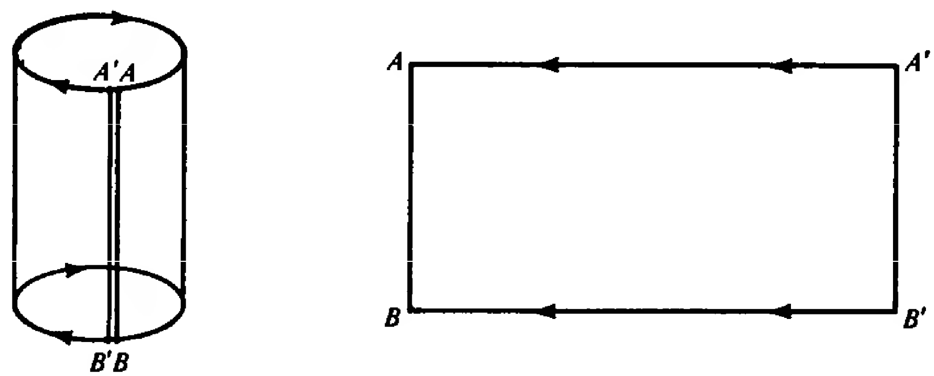

Figure 2.32

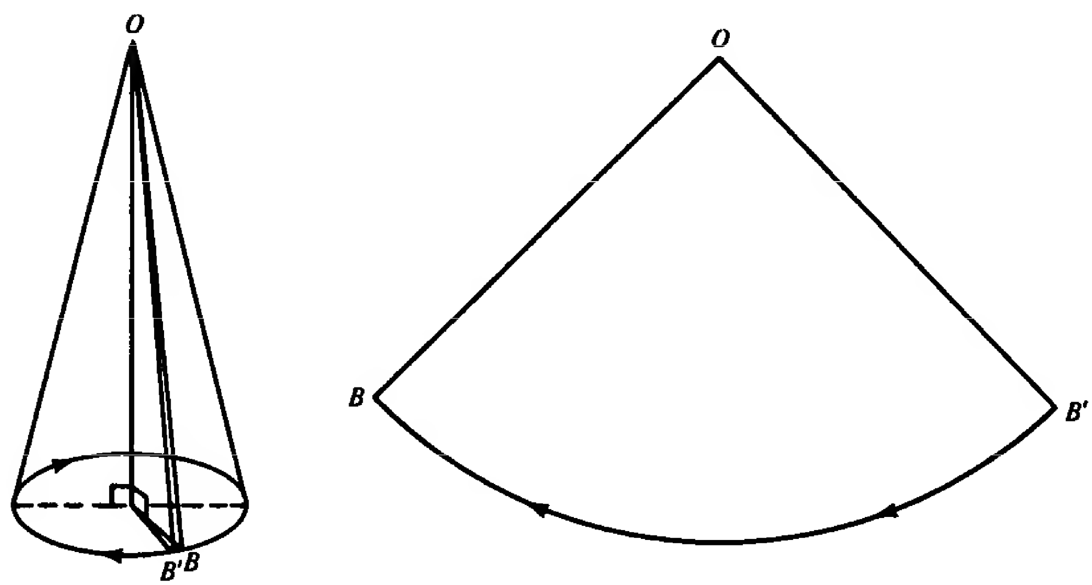

Figure 2.33

The second question, "What is the relevance of levers to determination of the volume of a sphere?", cannot be answered immediately. Every problem is viewed relative to a framework of associated ideas. First we must ask ourselves: What, for Archimedes, was the context of his problem? How did Archimedes, his mathematical contemporaries, and their mathematical predecessors conceive volume?

Well, of what volumes do you know the formulae? Certain volumes are easy to find; for instance, that of a rectangular parallelepiped—in brief, a

box. Its volume is the product of its length, breadth, and height. And do you know the formula for a right prism? But first, what is a right prism? It is a solid with a (plane) polygonal top congruent and parallel to its base, and lateral faces (parallelograms) perpendicular to its base. Its volume is the product of its base area and height. Note that a box is a prism (with a rectangular base) and therefore the formula for prisms is applicable. A right cylinder is a solid closely related to the prism. Is it not visibly evident that as the number of sides of the base of a right prism whose base is a regular polygon is increased, the prism approximates more closely a right cylinder? Doesn't this suggest that the same formula applies to right cylinders?

A pyramid presents a much more difficult problem. The formula for its volume is one-third area of its base times its height. Here, likewise, it is visibly evident that as the number of sides of a pyramid with a regular polygon as base is increased, the pyramid approximates more closely a cone. Surely we must anticipate that the formula for pyramids applies to cones.

Historically, who discovered the volume of the box, the prism, and the cylinder, it is impossible to say. But these formulae—and even those for the pyramid and cone—were already known to the Egyptians. They had no strict proofs, but the formulae were known and were used. There is, however, something definite to be said about the cone.

The volume of the cone was discovered by Democritus, who lived about 400 B.C. He did not prove it, he guessed it: the evidence is that his guess was not a blind guess, rather it was a reasoned conjecture. As Archimedes has remarked, great credit is due to Democritus for his conjecture since this made proof much easier. Eudoxus (408–355 B.C.), a pupil of Plato, subsequently gave a rigorous proof. Surely the labor of writing limited his manuscript to a few copies; none has survived. In those days editions did not run to thousands or hundreds of thousands of copies as modern books—especially, bad books—do. However, the substance of most of what he wrote is nevertheless available to us. Euclid, who lived about 300 B.C., wrote, to the knowledge of every schoolboy of my generation, *The Elements of Geometry*. Euclid's great achievement was the systematization of the works of his predecessors. His compilation includes quite a lot of things besides geometry; the Greeks understood the term in a more generous sense. The *Elements* preserve several of Eudoxus' proofs.

Archimedes studied and pondered deeply the works of his predecessors; these are the context within which he conceived the problem of the sphere. Herein lies the clue of the relevance of levers to volumes.

To find this clue we go back to Democritus. If you have heard his name before, it is more likely that you heard it in a philosophy lecture than in a

mathematics course. He is much better known as a philosopher, as an originator of atomic theory. Democritus' conception of an atom was something altogether different from today's physicists'. For him, as for the modern physicists, the whole world consists of atoms despite the apparent continuity of matter. The crucial difference is that the atom as Democritus conceived it could not be split. Matter could, conceptually at any rate, be chopped up into little bits, the little bits into smaller bits, until finally atoms were obtained; these little bits were held to be the smallest possible —indivisibles: one could chop no smaller.

It is worthwhile to stop for a minute or two to ponder how Democritus conjectured the volume of the cone. What can be said is necessarily, yet not solely, speculative; there are a few extant quotations to support us.

First, reconsider the volume of a rectangular parallelepiped, or box. How would we demonstrate to a child that a block whose edges are 3, 4, and 5 inches has a volume of 60 cubic inches? Surely, by splitting it into 60 cubes of unit edge and by pointing out that we agree to consider the volume of any such cube to be unit volume. We would, in short, consider the unit cube to be an atom and show that the block in question is made up of 60 atoms. The only objection is that by using *atom* in Democritus' sense we would be implying that a unit cube cannot be split into smaller cubes: a view to which we need not commit ourselves to effect our demonstration.

Next, could we demonstrate in exactly the same way that a block $3\frac{1}{2}'' \times 4'' \times 5''$ has a volume of 70 cubic inches? Not in exactly the same way since an edge of $3\frac{1}{2}''$ cannot form an edge of an integral number of unit cubes. The necessary modification is obvious. We chop the block into 560 cubes or atoms of $\frac{1}{2}''$ edge and then reassemble them to form 70 unit cubes. Thus the problem of determining the volume of a solid boils down to counting the number of its constituent atoms. Surely this was Democritus' basic idea.

Oh yes, the idea is simple, but the application can be arduous. Suppose that we are to demonstrate the volume of a block $3\frac{1}{10}'' \times 4\frac{1}{10}'' \times 5\frac{1}{10}''$. The counting of 64,821 atoms, cubes of $\frac{1}{10}''$ edge, is much quicker said than done. One counts up to 37,428, forgets whether that was the number of the atom just counted or the one about to be counted—and starts all over again. So, what do we do? We facilitate enumeration by dealing with large numbers of them *en bloc*—no pun intended. By multiplication we know immediately that a block with base 3 by 4 and height 5 has 12 atoms (unit cubes) in the first layer, 12 in the second layer, . . . , and so 5×12 atoms altogether. We enumerate the atoms by first dealing with one layer or cross section. Surely Democritus thought of this too.

What, for the cone, is a natural layer or cross section? Yes, a layer parallel to the base. So a cone is conceived as made up of adjacent

circular discs, just one atom thick. But here there is a complication. Although as for a rectangular block the successive layers or cross sections are all of the same shape, they are not all of the same size. The labor of enumeration would appear to force upon us the notion of a variable cross section.

Just how far Democritus developed this notion we do not know. Surely he knew that a cube can be dissected into three identical pyramids, so that the volume of a pyramid of this special shape is $\frac{1}{3}$ base $\times$ height; surely he must have conjectured that other pyramids and their limiting case, the cone, had the same formula.

Be this as it may, Eudoxus gave a rigorous proof. The proof is difficult and several lectures can profitably be given to its detailed exposition. The reader may try to read it in the twelfth book of Euclid's *Elements*.

Here we have the context, the conceptual background, of Archimedes problem to determine the volume of the sphere. A problem similar to, yet despite similarity, distinctly more difficult than, that of the cone. The sphere, unlike the cone, is rounded in all directions. His genius was equal to the challenge.

His method? A brilliant application of his Law (of Equilibrium) of the Lever. In accordance with his law he adjusted the length of the arms of a lever so that the cross section of a sphere counterbalanced both the corresponding cross section of a cylinder and the corresponding cross section of a cone. Simple? Although he called his method the *Mechanical Method*, he was no artificer of metals counterbalancing one chunk of sheet metal against a pair of chunks. He worked with ideas, not with tin; his method was conceptual. His corresponding cross sections were just one atom thick. Pray tell me, how thick is an atom? Oh no, it is thinner, much thinner than that. It's so small that if it were any smaller it would be no size at all. And what did Archimedes do? With an insolence to logic equaled only by the number of atoms he conceived, he inferred that the infinitely many cross-sections that fill the sphere would counterbalance both the infinitely many (corresponding) cross sections that fill the cylinder and the infinitely many (corresponding) cross sections that fill the cone. He inferred equilibrium of the solids from equilibrium of their cross sections. For full details the interested reader is referred to Vol. 1, pp. 155–158, of my book *Mathematics and Plausible Reasoning*, and to *Episodes from the Early History of Mathematics* by A. Aaboe, NML vol. 13, pp. 92–99.

I must add that Archimedes was too good a mathematician to leave it at that. He used the result obtained by his Mechanical Method only to *discover* the formula for the sphere: from discovery he proceeded to rigorous *proof*.

Also I must remark in passing that the notion of a variable cross section has had a long history. More than 2,000 years later we meet the idea of Cavalieri—in the terminology of Leibniz—of passing from infinitesimal element to integral whole: the idea of proceeding from integrand $f(x)dx$ to integral $\int f(x)dx$.

2.2.5 $(-)\cdot(-) = (+)$

Although Archimedes' discovery of the integral calculus is by far the most important application of his Law (of the Equilibrium) of the Lever, there is a multitude of other interesting applications. I shall conclude my selection from these with an answer to the perplexed schoolboy's question: Why does minus times minus equal plus?

One answer to this question is its proof; its detailed step-by-step deduction from the axioms and definitions of algebra. But this is not what your pupil is asking for. The proof, if presented, would go unappreciated; it demands sophistication beyond him.

His younger brother, who is learning to count, wants to know what $4 + 7$ is. His mother tells him, only to be driven crazy with his incessant "Why?", "Why?", "Why?". Is he demanding a résumé along the lines of Whitehead and Russell's *Principia Mathematica* proof that $1 + 1 = 2$? Or does he want the *comforting assurance* of a demonstration that four apples together with seven apples makes eleven apples, followed by similar demonstrations with oranges, his building bricks, and his mother's cups and saucers? Which answer does poor, distracted mother attempt?

Doubtlessly the "Why?" of your pupil, the older brother, though less incessant, is more demanding. His question may be many questions, "How was it discovered?", "What is its use?", yet the dominant demand of his "Why does minus times minus equal plus?" is for tangibility. It is no accidental figure of speech that we speak of *grasping* an idea; you must mother the brother. The lever meets his demand for tangible illustration.

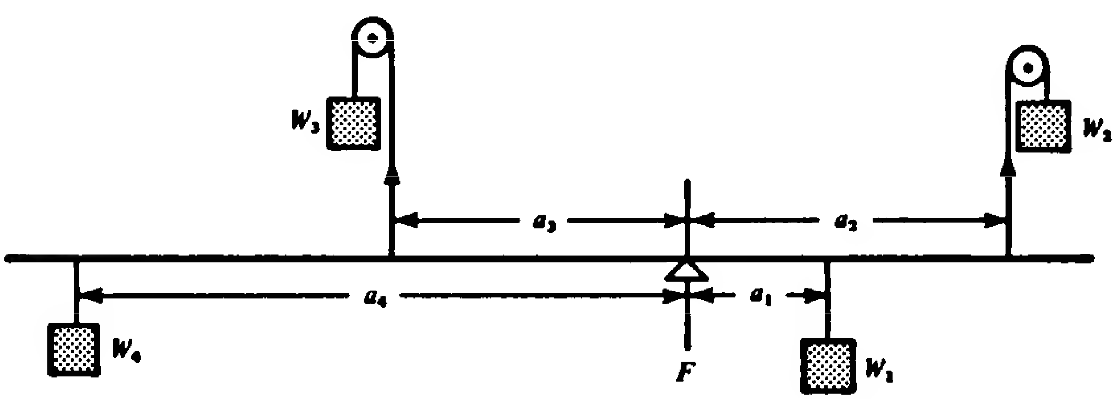

Figure 2.34

Consider the equilibrium of a (weightless) lever, acted upon by weights W_1, W_2, W_3, W_4 at distances a_1, a_2, a_3, a_4, respectively, from the fulcrum F, as illustrated by Fig. 2.34. Either a weight tends to rotate the lever about F in a clockwise direction $\circlearrowright$ (as do W_1 and W_3), or to rotate it in the opposite, anti-clockwise, direction $\circlearrowleft$ (as do W_2 and W_4). The measure of this tendency, the turning moment, is the product of the weight and the length of the arm from the fulcrum to the weight's point of application. More briefly,

$$\text{weight} \times \text{arm} \quad = \quad \text{moment}.$$

More precisely, this turning moment is termed *static* moment, in contradistinction to that considered in dynamics. Let us characterize a clockwise moment as positive and an anti-clockwise moment as negative.

On what does the characterization of a moment depend? Clearly it does not depend on the magnitude of the weight used; to increase W_1 is to increase the moment, not to alter its characterization. Nor does it depend upon the length of the arm; to shorten a_1 is to decrease the moment, not to change its sign. To alter the sign of the moment we must reverse the direction of the force due to (say) W by introducing a pulley, or hang it from the opposite side of the fulcrum. To take account of these pertinent considerations let us term a weight whose force acts vertically downwards from the lever (as do W_1 and W_4) a positive weight, and in contradistinction, a weight whose force acts vertically upwards from the lever, negative (as do W_2 and W_3). And to distinguish between a weight acting to the right of the fulcrum (as do W_1 and W_2) and a weight acting to the left (as do W_3 and W_4) we introduce an x-axis coincident with the lever, with origin at the fulcrum, so that each arm a is a horizontal, directed line segment. As in drawing graphs, we consider to the right from the origin to be positive and the opposite direction to be negative. Thus the arms a_1 and a_2 are positive; a_3 and a_4, negative.

Fig. 2.35 indicates the signs of the weights and the arms of Fig. 2.34 and the characterizations or signs of the corresponding moments. W_1 (positive) acting at a distance a_1 to the right of O (positive) has a tendency to give the lever a clockwise, positive, rotation; i.e., the product of a positive weight and a positive arm is a positive moment. And remembering the Cheshire cat who disappeared so hastily that he left his grin behind, we may put the matter schematically

$$+ \cdot + = +.$$

There is a superstition that mathematical notation must be always perfect. But English with never a colloquialism, never an ellipsis is

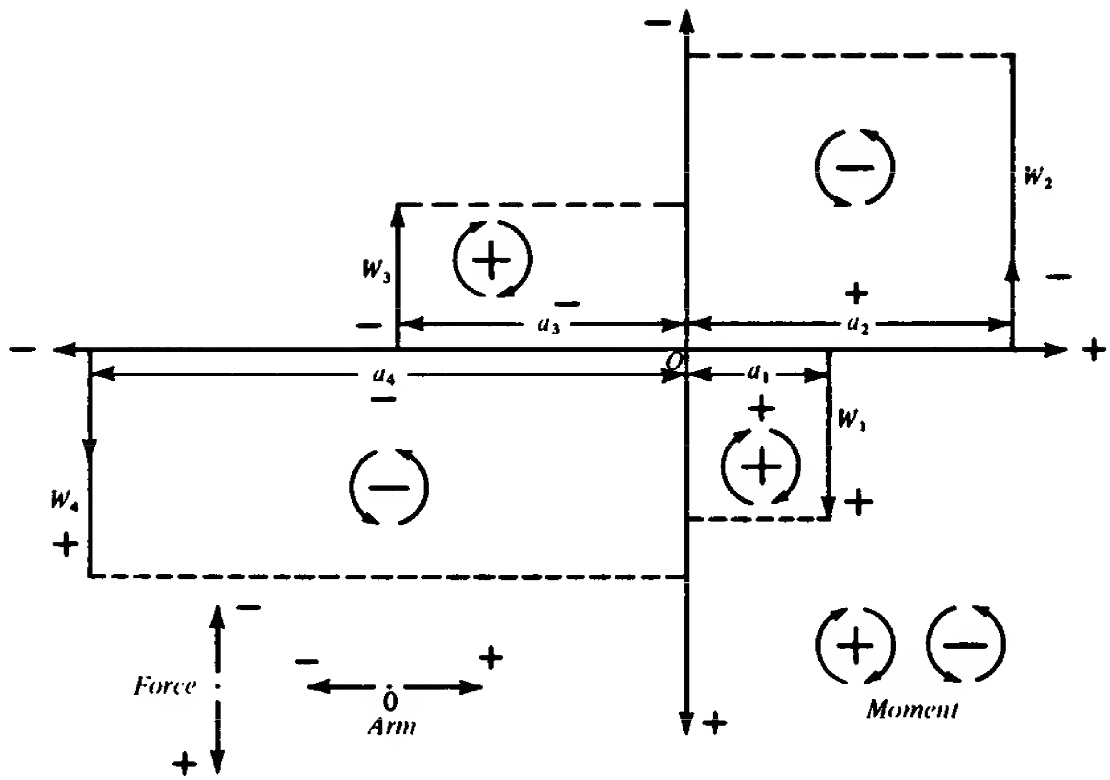

Figure 2.35

unendurable; it leaves the reader nothing to do but passively listen. As with English, so with mathematics. Let us consider the grins without the cats.

What is the sign of W_2? Yes, negative. And the sign of a_2? Yes, positive. And their moment? Yes, anti-clockwise. So,

$$- \cdot + = - .$$

It is left to the reader to satisfy himself that consideration of W_4 and a_4 gives

$$+ \cdot - = -$$

and of W_3 and a_3 gives

$$- \cdot - = + .$$

Have we proved it? No, we have not deduced it from the definitions and axioms of algebra. But, we have shown that it has an intuitive interpretation, that it is applicable to physics; above all, we have made it tangible. Of course, any physical phenomenon whose magnitude is the product of the magnitude of two physical quantities will serve to illustrate the rule provided that each magnitude is capable of taking both positive and negative signs. Yet what can be a more elementary, or a more

intuitive, illustration than that furnished by the lever? And is it now so very difficult to conjecture how that minus times minus equals plus was first discovered?

2.2.6 Von Mises' Flight Triangle

We have discussed instances of equilibrium, namely, inclined plane, pulley, and lever, in terms of vectors. Our next example, although strictly speaking a problem of dynamics, is so simple that we include it in our vector treatment of statics. Our problem is how to determine the air speed of an airplane.

First, what do we mean by *air speed*? We do not mean *ground speed*. The former is the speed of the plane relative to the air it flies through; the latter, its speed relative to the ground it flies over. This distinction is vital to our problem. To fix it clearly in our minds let us ponder the following illustration.

Suppose that an airplane, flying at constant speed, goes from San Francisco to Los Angeles in 4 hours. For easy arithmetic, let us take the distance to be 400 miles. It follows that the airplane's ground speed—or, if you prefer, road speed—is 100 m.p.h. That is, it gets there in the same time and at the same speed as a motor car would, if automobile engines were just a little more powerful, the San Francisco-Los Angeles road much less congested, and the California speed cops less vigilant. Clearly, an airplane that keeps pace with a car racing to Los Angeles at a road speed of 100 m.p.h. must itself have a road speed of 100 m.p.h.

But what is the air speed of the plane? That depends on the speed of the air. If the air is still, then the plane flies through the air at the same speed as it flies over the ground. Its air speed is 100 m.p.h., the same as its road speed.

Next suppose that the car stops for gas and that the plane overhead is battling against a 200 m.p.h. head-on hurricane. The plane, to continue to keep pace with the car, to remain directly above the filling station while the car refuels, must be flying through the hurricane at 200 m.p.h. Although the plane's road speed is now zero (as is the car's), its air speed, its speed through the air, is now 200 m.p.h. When the car, refueled, continues its journey at 100 m.p.h., the plane to keep pace with it must (because of the head-on hurricane) increase its air speed to 300 m.p.h. In short, the air speed of the plane is the road speed (or, if you prefer, ground speed) it would have if flying in still air.

Let us now use vectors to make visibly obvious the relation between the plane's road velocity $\vec{v}$ (i.e., a road speed v in the direction PL), its air velocity $\vec{a}$ (i.e., an air speed a in the direction PO'), and the wind's velocity $\vec{w}$ (i.e., air moving with a speed w in the direction PO). See

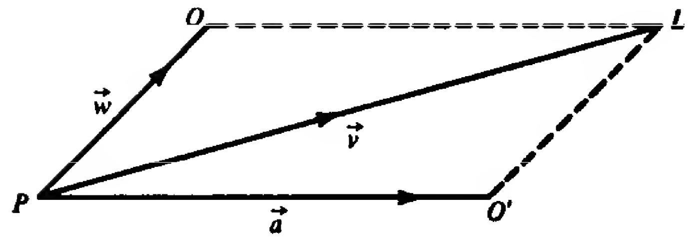

Figure 2.36

Fig. 2.36. The situation is analogous to that, considered earlier, of a motorboat crossing a river. With no wind, the plane would, say in one minute, fly from P to O'. With no air speed, a balloon would, in the same time, drift with the wind from P to O. And we have already seen that the resultant of simultaneous displacements is as if they had been consecutive. Thus, the actual path of the plane is PL; its road velocity $\vec{v}$ is the resultant of its air velocity $\vec{a}$ and the wind's velocity $\vec{w}$. As we anticipated, when there is no wind (so that $\overrightarrow{PO}$ is zero and L and O' coincide) the actual road velocity $\overrightarrow{PL}$ and the air velocity $\overrightarrow{PO'}$ of the plane, are identical.

So, to determine the air speed of a plane it is sufficient merely to determine its maximum road speed on a calm day. The snag is the sparsity of windless days. Airplane manufacturers want to make money as well as planes, and so cannot afford to sit around for six months waiting for still air over some hundreds of square miles in which to test the performance of their machines. Impatiently, you exclaim, "Why wait for a windless day?" True if $\vec{w}$ can be accurately measured, as well as $\vec{v}$, then $\vec{a}$ is readily calculated by means of a vector parallelogram of velocities. The snag, here, is to measure $\vec{w}$ accurately. The practical problem is how to determine, from the road speed of a plane at full throttle in a wind *whose velocity* $\vec{w}$ *is not known*, the maximum road speed of the plane when there is no wind. Quite a problem. It was solved by Von Mises some fifty years ago; this I well remember as I heard it from him at that time.

We are now ready to begin introducing his method. An airplane flies at full throttle from A along a triangular course ABC, a flight triangle, whose vertices are chosen to be easily identifiable landmarks, preferably a few hundred miles from one another. The lengths of the legs AB, BC, CA, being known, and the time to traverse each one of them recorded, the road speed for each is readily calculated. But of course the directions of the legs are also known, so that we know the three velocity vectors, say, $\vec{v}_1$, $\vec{v}_2$, $\vec{v}_3$. See Fig. 2.37. And since the plane was flown at full throttle, its air speed a is its maximum air speed. How is a to be determined from the data?

It is important to note that the sides of the flight triangle represent the directions of the velocity vectors, but not, except when there is no wind at all, their magnitudes. We use heavy arrows in Fig. 2.37 as a graphical device to indicate the road velocity in direction *and* magnitude. Suppose, for example, that in flying the longest leg BC the plane has a head wind. In consequence, in flying BC the airplane's road speed is slower than on either of the other two legs—assuming, of course, that the wind remains constant in both speed and direction. But, if the least road speed is represented by the length of the longest side, the greater road speeds would need be represented by sides longer than the longest. Agreed?

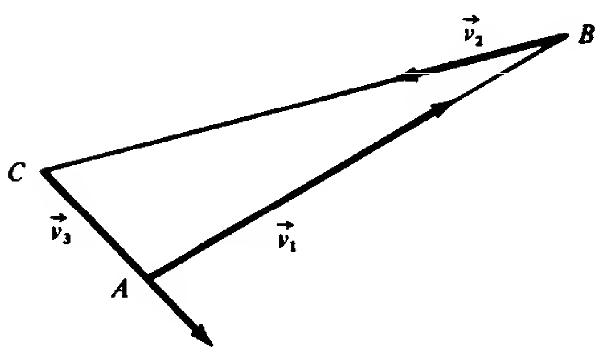

Figure 2.37

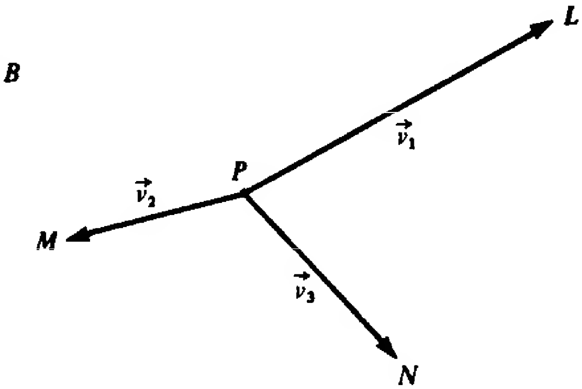

Figure 2.38

To make the data fully visible, starting from some point P, we draw line segments PL, PM, PN, to represent $\vec{v}_1, \vec{v}_2, \vec{v}_3$, respectively, three vectors (in Fig. 2.38) which agree with the three arrows in Fig. 2.37 in magnitude as well as in direction, specifically $PL \| AB$, $PM \| BC$, $PN \| CA$.

How are we to utilize this vector diagram to determine a, the maximum air speed of the plane? Can we introduce a vector for $\vec{a}$? Alas, no. That the plane flies around the flight triangle at full throttle merely implies that its air speed a, the magnitude of $\vec{a}$, is constant. It does not imply that the direction of $\vec{a}$ is constant. We must think again.

Can we introduce a vector $\vec{w}$ for the wind? Doesn't this suggestion seem more fruitful? Remember our assumption that the wind velocity remains constant throughout the entire flight, so that one directed line segment from P should serve as a vector component of the road velocity along *each and all* of the three legs of the flight triangle. As in Fig. 2.36, let PO be a directed line segment representing $\vec{w}$. But wait a moment; $\vec{w}$, although constant, is unknown. Are we baffled? Think a moment. Do we, in algebra, delay embodying an unknown, x, in equations until we have determined its value? No, to the contrary, we put in x in order to determine its value. So? Tentatively, we insert O. See Fig. 2.39.* What

* We shall determine its precise location in a moment.

now? Reference to Fig. 2.36 suggests completion of the parallelogram of Fig. 2.39 of which PO is a side and PL a diagonal. However, in Fig. 2.36, OL is parallel and equal to PO', an equivalent directed line segment, so that $\vec{a}$ could alternatively be represented by OL and the vector parallelogram dispensed with in favor of the vector triangle POL. Thus, in Fig. 2.39, $\overrightarrow{OL}$ will be an air-speed vector which together with $\vec{w}$ has the resultant $\vec{v}_1$. And what of triangles POM, PON? We have remarked that $\overrightarrow{PO}$ should serve as a component of the road velocity along each leg of the flight triangle. See Fig. 2.40.

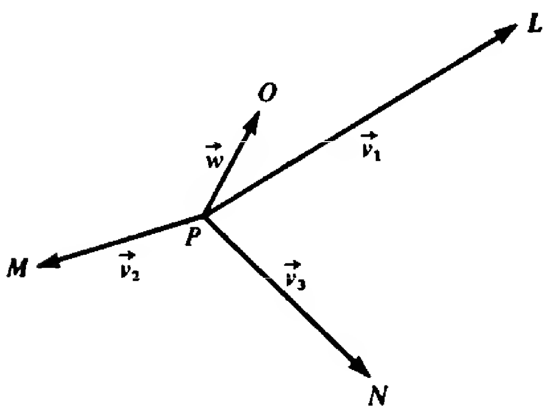

Figure 2.39

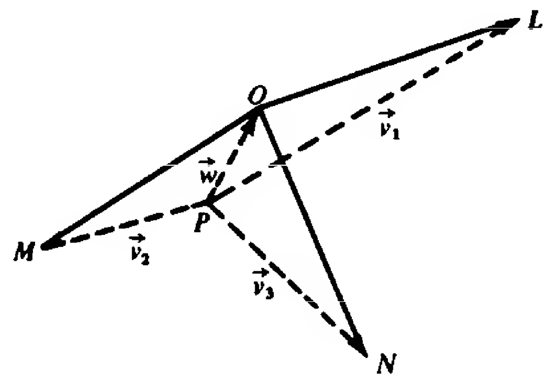

Figure 2.40

Note how delightfully the three vector triangles "interlock" on PO just like pieces of a jigsaw puzzle. Not only is $\vec{v}_1$ the resultant of $\vec{w}$ and an air-speed velocity $\overrightarrow{OL}$; also $\vec{v}_2$ is the resultant of $\vec{w}$ and an air-speed velocity $\overrightarrow{OM}$, and $\vec{v}_3$ the resultant of $\vec{w}$ and an air-speed velocity $\overrightarrow{ON}$. But the plane flew the entire flight at full throttle. Therefore its air speed a (but not its air velocity) was the same for each leg. Therefore $\overrightarrow{OL}$, $\overrightarrow{OM}$, $\overrightarrow{ON}$ have the same magnitude (but not same direction). Therefore O is equidistant from L, M, N; it must be the center of the circumscribing circle of triangle LMN. Construction of O determines both a and $\vec{w}$. This is Von Mises' elegant, ingenious solution.

CHAPTER THREE

From the History of Dynamics

Whereas statics, we recall, is that part of mechanics which is concerned with the equilibrium of bodies, dynamics is that part which is concerned with the motion of bodies. The former, as we have had occasion to note, goes back to the Greeks; to Archimedes' discovery of the Law of the Lever and his application of it to the integral calculus. The latter is relatively new; it starts with Galileo.

SECTION 1. GALILEO

Galileo is known by his first name; his family name is Galilei. He was born in 1564 and died in 1642. To believers in the transmigration of souls the date of his death is important. Not only did he die in the year in which Newton was born, conveniently for their speculations, he died shortly before Newton was born. A much more important date is 1636, the year in which he completed the book on which his fame so securely rests, the *Dialogue Concerning Two New Sciences*. Although many of his brilliant predecessors, beginning with Aristotle, and including that most versatile of versatile geniuses, Leonardo da Vinci, had been interested in the free fall of heavy bodies, Galileo was incomparably the greatest dynamicist of them all. He inherited a dogma and bequeathed a science.

His tomb is to be found in Florence, in the Church of Santa Croce, among those of Leonardo and Michelangelo the artists, Dante the poet, and Machiavelli the politician. His instruments are also to be found in Florence, in the Museum of the History of Science; among them the telescopes he made, used, but did not invent, and the thermometers he made, used, and did indeed invent; also his instruments for the study of dynamics. Florence is an interesting city.

His father sent Galileo to study medicine, but he was soon bored by it. In those days the college course was a digesting, and the examination a regurgitation, of the texts of Galen. Galen had lived from about 130–200 A.D. Meanwhile, his texts—presumably in Latin, for then as now few knew Greek—had been accumulating the dust of dogma for fourteen centuries. In medicine it was sufficient to quote Galen, as to quote Aristotle in practically everything else. For Galileo, to quote was not sufficient; he turned to mathematics.

3.1.1 Heavier Bodies Fall Faster?

Aristotle had stressed the importance of observation, yet he did not, in dynamics, observe well himself. It is a matter of casual observation, well known to mountaineers and others, that bodies free to fall, fall to the ground. To Aristotle's very casual observation, the heavier the body the faster the fall; Galileo argues to the contrary.

Suppose two bodies W, w to fall freely from rest and to have velocities V, v, respectively, at the end of unit time. Then, according to Aristotle, given that W is greater than w, V must be greater than v. But, asks Galileo, "What happens if the two bodies are conjoined?" Let U be the velocity at the end of unit time of the conjoint body $W + w$. Since w alone falls more slowly than W, the w part of $W + w$ must retard the W part; the fleet of foot has to slow down to help the lame along. U must be less than V. Yet, since $W + w$ is greater than W, by Aristotle's hypothesis, U must be greater than V. Therefore U is both less than and greater than V; $W + w$ falls both slower than and faster than W. This is absurd.

What has Galileo done? He has said, in effect, that here is a possible law, supported by a rather weak observation. Is it consistent? He has argued that it is not. Therefore it is unacceptable to mathematics; it cannot be an ingredient of a systematic description of phenomena.

Galileo's argument was an important one; it made uneasy the dogmatic slumbers of many of his contemporaries. He spoke and wrote with an edge to tongue and pen. Like his father, he was quarrelsome as well as argumentative, and witty as well as logical—a combination that made his opponents look silly and their arguments unsound. He did not endear himself to all.

3.1.2 Not "Why?", But "How?"

Why? Why this? Why that? Such are the questions asked by the good shepherd Aristotle and bleated by his sheep down through the centuries. Why do heavy bodies fall? "Because," says Aristotle, "each body seeks its natural place." He argues as if an inanimate object were an animal

seeking its mate. Are you much enlightened by this argument? No, because you are born in modern times; Galileo was not. He had to argue the point; such was the intellectual climate of his day. Galileo, frighteningly modern, asked a better question; not "Why?", but "How?". His question was a demand for precise description of the phenomenon under consideration, not speculative anthropomorphism. "How," he asked, "do bodies fall freely?" His "How?" was much more. Behind his question stood his fundamental tenet: The great book of Nature is written in mathematical language. (See the motto on front cover.) His demand was for a precise mathematical law, no less.

3.1.3 How Do Heavy Bodies Fall?

Galileo asked the right kind of question. Finally he asked the right question of the right kind. He gave the right answer. In so doing, he founded a new science.

How do heavy bodies fall? The farther, the faster. Even the most casual observer cannot avoid the conclusion that the velocity increases with the distance fallen. We all know that the greater the drop of a hammer, the harder it hits. We all know that it is better to be hit on the head by a bag of bone meal dropped from the second story window than by one dropped from the top of the Empire State Building. It can be made painfully evident that free-fall motion is accelerated. So what is the mathematical law relating the velocity to the distance? What is the simplest conjecture? That the velocity of the falling body is directly proportional to the distance fallen? This was Galileo's first question. It is the right sort of question.

It is likely that Leonardo and a few others before Galileo had raised some similar question; the difference is that Galileo took it more seriously. After some years of pondering he came to the conclusion that this conjecture is absolutely untenable; it has the self-contradictory consequence that the free fall could never get started. He reached this conclusion by a highly ingenious and subtle elementary argument which deserves our admiration, yet we shall not discuss it here.* With present day mathematical technique we can give a quite straightforward argument by using a differential equation, as we shall see later (Number 5.1.2). So, Galileo's conjecture, based on the undeniable fact of free fall that the farther the fall, the faster the fall, is untenable. The fall could never get started. He had to think again. But, it is also an unavoidable observation that the longer the time, the faster the fall. The bone meal takes longer to hit you the harder when dropped from the top of the Empire State

*See Pólya, *Mathematics and Plausible Reasoning*, Vol. 2 (2nd edition, 1968), pp. 207–209.

Building instead of being tossed from a second-floor window. Free fall is accelerated with respect to time as well as with respect to distance. So what is the mathematical law relating velocity to time? What is the simplest conjecture? That the velocity of the falling body is directly proportional to the time? That is Galileo's second question. It turned out to be the right question of the right kind.

How did Galileo verify his conjecture experimentally? Remember he did not have today's elaborate photoelectric equipment with which to handle split-second motion. Stop a moving particle to take a longer look and you have destroyed the velocity you wished to observe. Yet there is no hurry in measuring distance; this can be done at leisure and with accuracy. So Galileo's problem was to *deduce* the law relating distance to time *implied by* (*and implying*) velocity being directly proportional to time, and hence indirectly to verify the latter relation by direct verification of the former.

So, to understand fully Galileo's indirect experimental verification of his conjecture, we must first ask how he deduced that relation between distance and time which is implied by his conjecture. To facilitate deduction, like Galileo but unlike Aristotle, we shall use diagrams with coordinate axes; "essential characters of mathematical language," as Galileo would put it. Galileo's investigation of dynamics was physical; Aristotle's was metaphysical. But, unlike Galileo, we have the additional convenience of algebraic notation. Had it been invented in his day he would certainly have known it; almost certainly he would have been able to push his development of dynamics much farther.

Suppose that a heavy body has a velocity v when it has been falling freely for time t. Then Galileo's hypothesis is that v is directly proportional to t; that v is a constant multiple of t; that

$$v = \text{constant} \times t.$$

The numerical value of the constant depends upon the units we use for v and t, and the constant is nowadays usually denoted by g. Our primary school teacher who taught us that A is for apple, should have added that g is for gravity that made the apple in Newton's orchard fall. Thus, algebraically speaking, Galileo's conjecture is

$$v = g \cdot t. \tag{1}$$

But the distance s fallen from rest in time t by the heavy body depends upon t; s is a specific, yet unspecified, function of t;

$$s = f(t). \tag{2}$$

Galileo's problem is: Given (1), to specify (2).

How did he solve it? Most ingeniously, by conceiving accelerated, non-uniform motion as a limiting case of non-accelerated, uniform motions.

First, consider uniform motion. If you drive for two hours at a steady rate of 40 m.p.h., you go a total distance of 80 miles.

$$80 = 40 \times 2.$$

More generally,

$$\text{distance} = \text{uniform velocity} \times \text{time}.$$

Algebraically,

$$s = v \times t, \tag{3}$$

where v is constant. Graphically, see Fig. 3.1. The ordinate v is constant, so that the graph of v is a straight line parallel to the t-axis. Note that the area under the curve, i.e., the area of the shaded rectangle, is $v \times t$. So, by (3), the total distance traveled is represented by the area under the curve—when the motion is uniform. Oh yes, an obvious observation, but nevertheless important.

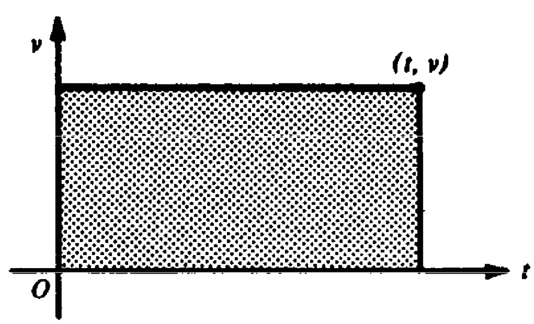

Figure 3.1

Second, consider the non-uniform motion. What is the graph of (1)? This equation is of the form $y = mx$, with v instead of y, t instead of x, and g instead of m. It is a straight line through the origin with slope g. See Fig. 3.2.

Why is the velocity of a freely falling body not uniform? Because its velocity continually increases, of course. And an accelerating car does not, for example, move at 0 ft/sec for the first second, at 5 ft/sec for the second second, at 10 ft/sec for the third second, at 15 ft/sec for the fourth second, and so on. To the contrary, let us temporarily suppose

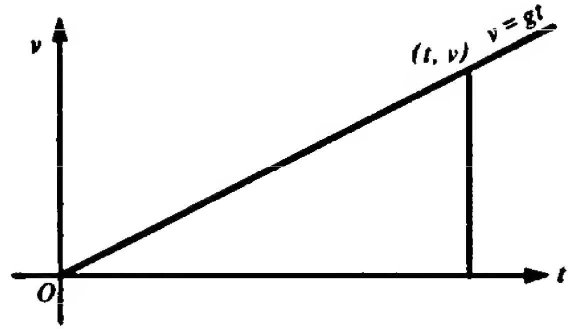

Figure 3.2

velocity which continually increases at a steady rate to be characterized by such spasms of uniform motion punctuated by accelerating jerks at the end of regular intervals. This grotesque caricature of the truth is illustrated by Fig. 3.3. In the first second the car moves zero feet, then instantaneously with a bone-shattering jerk it accelerates to 5 ft/sec. After a second of gentle driving at this constant velocity there is a we-had-better-install-safety-belts jerk to 10 ft/sec. There follows a second's driving at 10 ft/sec; jerk; a second's driving at 15 ft/sec; jerk; a second's driving at 20 ft/sec; and so on. The distances covered in the successive intervals are represented by the areas of the successive rectangles. (The first rectangle is of zero height.) The total distance traveled is represented by the total area of the shaded rectangles.

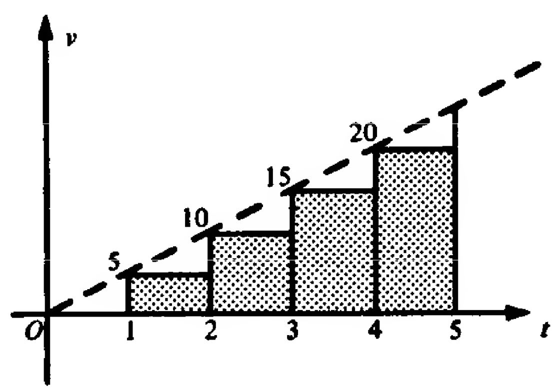

Figure 3.3

Now suppose the accelerations and the time intervals to be halved. In the first five seconds the car now acquires successively, for half-second intervals, the ten velocities, $0, 2\frac{1}{2}, 5, 7\frac{1}{2}, 10, \ldots, 20, 22\frac{1}{2}$ ft/sec. Illustrate this for yourself by a diagram of the same type as Fig. 3.3. The jerks, although twice as frequent, are only half as strenuous, for the sudden increases in velocity are now only $\frac{5}{2}$ ft/sec instead of 5 ft/sec.

Now suppose these accelerations and time intervals to be halved too. Although the jerks are four times as frequent as in the initial case, they

are only one-quarter as strenuous; the sudden increases in velocity are now only 5/4 ft/sec instead of 5 ft/sec. With the jerks eight times as frequent they are only one-eighth as strenuous; the sudden increases in velocity are now only 5/8 ft/sec instead of 5 ft/sec. When the intervals are each $1/2^n$ of a second where n is large, the jerks become gentle jerks, for the sudden changes in velocity have been decreased to $5/2^n$ ft/sec. The larger we make n, the more nearly we smooth out our ride. By making n sufficiently large we make the smoothness of our ride differ imperceptibly from the glide of a car whose velocity is continually increasing at a steady rate. By making n sufficiently large, our grotesque caricature becomes a description of continually increasing velocity as close to the real thing as we please.

Mutatis mutandis, these considerations of course apply equally well to freely falling bodies. And what happens to Fig. 3.3 and the figure that you have drawn for yourself when n becomes large? As the rectangles become more numerous they wear a leaner look and more completely fill the area under the curve of $v = g \cdot t$. By making n sufficiently large we come arbitrarily close to filling the whole area. See Fig. 3.4. So? Why, of course, the area under the curve in Fig. 3.2 (i.e., the shaded area in Fig. 3.4 (b)) represents the total distance traveled in time t by a heavy body falling from rest. Despite the fact that the motion is non-uniform, the total distance traveled is, as in the case of uniform motion (illustrated by Fig. 3.1), nevertheless represented by the area under the curve. But, the area under the curve is a triangular area of base t and height gt. So,

$$s = \tfrac{1}{2} t \times gt$$

i.e.,

$$s = \tfrac{1}{2} gt^2. \tag{4}$$

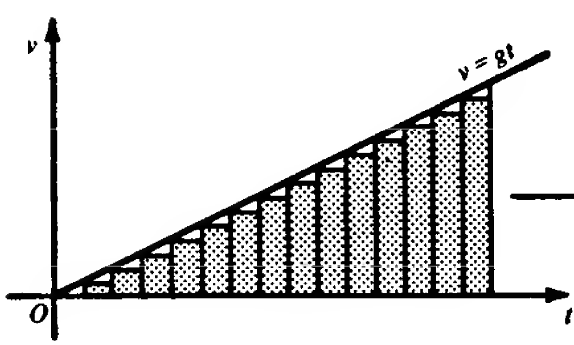

Figure 3.4(a)

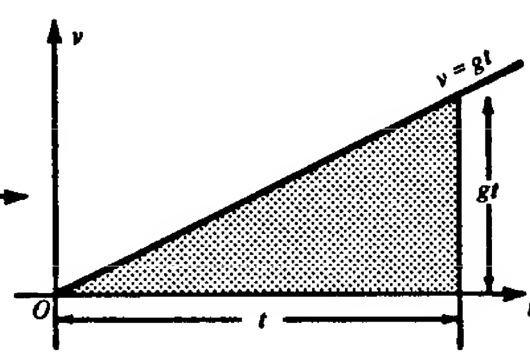

Figure 3.4(b)

This is the way in which Galileo deduced the specification of $f(t)$ of equation (2). This is the law relating distance to time implied by velocity being directly proportional to time. (And is it not evident that if the area under the curve is $\frac{1}{2}gt^2$ for all values of t, then the equation of the curve must be $v = gt$?) The distance fallen is proportional not, as Galileo first thought, to the time, but to the square of the time. In disproving the former and deducing the latter, he investigated two important corners of the calculus.

Galileo's basic difficulty, we recall, was that he could not "freeze" the motion of a falling body to take a longer look at its instantaneous velocity; his guiding motive, that distances are easier to measure than velocities. His final problem was to verify (4) experimentally, thereby verifying indirectly its implication, (1). How did he do this?

Consider the following tabulation.

t	Total distance fallen in t seconds $= \frac{1}{2}gt^2$	Distance fallen in successive seconds
0	0	
1	$\frac{1}{2}g \cdot 1$	$\frac{1}{2}g \cdot 1$
2	$\frac{1}{2}g \cdot 4$	$\frac{1}{2}g \cdot 3$
3	$\frac{1}{2}g \cdot 9$	$\frac{1}{2}g \cdot 5$
4	$\frac{1}{2}g \cdot 16$	$\frac{1}{2}g \cdot 7$
5	$\frac{1}{2}g \cdot 25$	$\frac{1}{2}g \cdot 9$

The distances fallen in successive equal time intervals are in the ratio 1:3:5:7:9 and so on.

Thus if a heavy body dropped from the top of a wall passes a chalk mark 1 unit down at the end of one second, it should pass a mark 3 units farther down at the end of two seconds, a mark 5 units farther down at the end of three seconds, and so on. But bodies fall so fast that even these observations are difficult; despite a legend to the contrary Galileo did not drop cannon balls from the Leaning Tower of Pisa. Isn't it possible to slow up the motion to facilitate observation? A reduction in the value of $\frac{1}{2}g$ would not alter the ratios. A vertical wall is a limiting case of an inclined plane; shouldn't we expect these ratios to hold for motion on an incline? Unlike a vertical, an incline takes some of the weight of the body sliding along its surface, thereby reducing the body's acceleration. Surely the smaller the angle of inclination α, the slower the motion.

Galileo experimented to find out. See Fig. 3.5. He found that a ball let roll from O, which moved from O to A in unit time, moved from B to C, and from C to D, also in unit time. As near as he could tell, this phenomenon was independent of the angle of inclination of the incline. In

this way Galileo substantiated his right answer to the right question. (See the last paragraph of Number 3.1.7.)

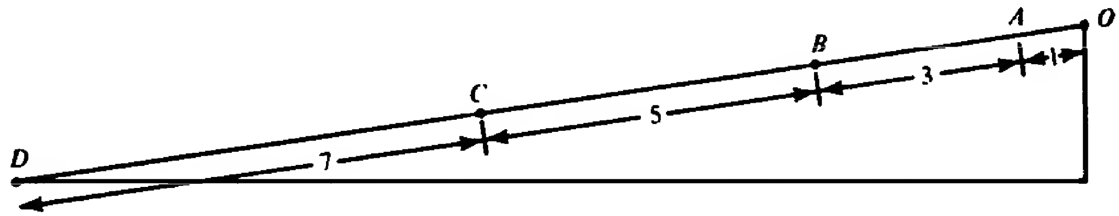

Figure 3.5

3.1.4 Dynamics of the Inclined Plane

When the angle of inclination α is reduced to zero, the plane is horizontal and a body on it does not move; the plane takes the whole of the weight. The greater the angle α, the faster the body slides down and the smaller the proportion of its weight taken by the plane. Finally, when $\alpha = 90°$, the plane takes none of the weight; we have free fall. Obviously the proportion of the weight taken by the plane depends on α as does the condition for equilibrium of a body on it. The condition for the latter, which Galileo knew either from Stevinus or by figuring it out for himself, helped him to deduce the former. His method is what really amounts to an implicit use of a parallelogram of forces.

First, what causes a body to accelerate? Yes, the force acting on it. We all know that to speed up when driving, to accelerate, we have "to step on the gas," as we say. Our engine has to deliver more force. And what is the force which causes a freely falling body to continually increase its velocity? Yes, the gravitational pull of the Earth, its weight. We now know what Galileo could not know, that the acceleration of the free fall of a body to the Moon's surface is only about one-sixth that to the Earth's surface. Although the substance of the body is unchanged in moving it from the Moon to the Earth, its weight is increased about sixfold. On the Moon it weighs less because it is in a weaker gravitational field. There, with only one-sixth the effort to surmount an overhang, rock climbing must be a less strenuous affair. And when you slip and fall off you have only one-sixth the terrestial acceleration. The free fall of a body—its acceleration—is proportional to the force acting on it—its weight.

Next, what is the g of equation (1)? Consider Fig. 3.2, the graph of this equation. What is m in $y = mx$? Yes, m is the slope. More explicitly, m is the ratio of the change in vertical displacement to the change in horizontal displacement. So, *mutatis mutandis*, g is the ratio of the increase in velocity to the increase in time. But the curve is a straight line, a curve of constant slope, so that the ratio g is the same no matter how small the change in time. In short, g is the constant rate of

instantaneous change of velocity due to gravity—in a word, the acceleration.

To sum up: g of equation (1) is the gravitational constant, the measure of the Earth's gravitational field. The force exerted on a body by the Earth's, the Moon's, or any other gravitational field, is proportional to the constant for that field.

Accordingly, let us take g as a measure of the force acting vertically downward on a body on an inclined plane. See Fig. 3.6. Since the surface of the incline is supposed to be perfectly smooth, the only effect of the plane's reaction R, must be perpendicular to its surface. But, recalling the geometry of Number 2.2.1 (Vectors, Inclined Plane) the vector g may be resolved into a force $g \cos \alpha$ perpendicular to the plane (and so equal and opposite to R as there is no motion perpendicular to the incline) and a force $g \sin \alpha$ down the incline. Thus, the problem of free motion down a smooth incline becomes, in effect, that of a body "falling" in a gravitational field of $g \sin \alpha$ (instead of g) which acts in the direction OA (instead of vertically downwards).

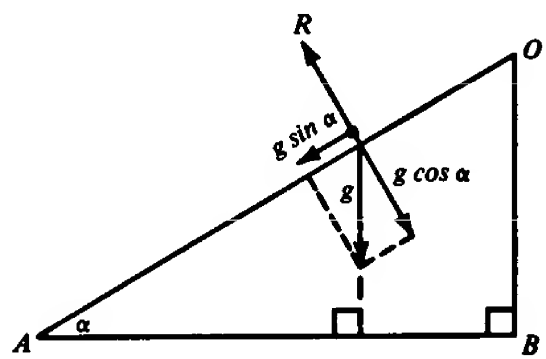

Figure 3.6

The gravitational field being $g \sin \alpha$ instead of g, instead of

(1) $$v = g \cdot t$$

we have

(1′) $$v = g \sin \alpha \cdot t.$$

But,

(4) $$s = g \cdot \tfrac{1}{2} t^2$$

is a consequence of (1), so that

(4′) $$s = g \sin \alpha \cdot \tfrac{1}{2} t^2$$

is a consequence of (1′). Taking $t = 0, 1, 2, 3, \ldots$, it follows, as we anticipated, that displacements down the plane in consecutive unit intervals are in the ratio 1:3:5:7: . . . .

This completes our exposition of Galileo's deduction of (4′). His treatment was much less explicit.

From the relation (1) between v and t for free fall Galileo deduced, as we have seen, the relation (4) between s and t. He went on to ask what is the relation between v and s. The answer to this question is the elimination of t from (1) and (4). Dividing (1) by g and squaring, we have

$$\frac{v^2}{g^2} = t^2.$$

Substituting for t^2 in (4), we get

$$s = \frac{1}{2} g \cdot \frac{v^2}{g^2} = \frac{v^2}{2g},$$

so that

$$v^2 = 2gs. \tag{5}$$

Next, he asked the same question for free motion down an inclined plane of angle α. Remembering that (1′) is similar to (1) and (4′) to (4) in that the latter pair are the former pair except for the factor $\sin \alpha$, what do you anticipate for motion down the incline? Do you not expect an equation (5′) which has the same similarity to (5) as (1′) has to (1) and (4′) to (4)? Yes, we conjecture,

$$v^2 = 2g \sin \alpha \cdot s. \tag{5′?}$$

We have committed ourselves; we must test our conjecture.

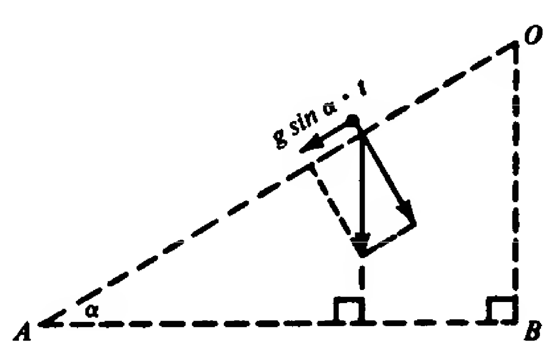

Figure 3.7

Referring to Fig. 3.7, we suppose a body which starts from rest at O, i.e., with $v = 0$, $s = 0$, and $t = 0$, to reach A, the bottom of the incline, with velocity V, after traveling the distance S from O to A, in time T. By (1′)

$$V = g \sin \alpha \cdot T,$$

and by (4′)

$$S = \tfrac{1}{2} g \sin \alpha \cdot T^2.$$

Dividing the former by $g \sin \alpha$ and squaring, we have

$$\frac{V^2}{g^2 \sin^2 \alpha} = T^2,$$

and substituting for T^2 in the latter,

$$S = \frac{1}{2} g \sin \alpha \cdot \frac{V^2}{g^2 \sin^2 \alpha} = \frac{V^2}{2g \sin \alpha},$$

so that

(5′) $$V^2 = 2g \sin \alpha \cdot S.$$

This, with appropriate change of notation, is (5′?). Our conjecture is confirmed.

(5′) has a vitally important consequence for the entire development of dynamics. In Fig. 3.7, let OB, the total vertical drop during the free motion along OA, be H. Then, since $OA = S$, we have

$$\sin \alpha = \frac{H}{S}.$$

Substituting for $\sin \alpha$ in (5′), we obtain

$$V^2 = 2g \cdot \frac{H}{S} \cdot S,$$

so

(6) $$V^2 = 2gH.$$

Hasn't Galileo's question a truly astonishing answer? (6) makes no reference to the length of the incline nor to its angle of inclination. The

square of the velocity—and consequently the velocity itself—is independent of these things. The velocity acquired depends solely upon the height lost from the commencement of the motion. And since the acquired velocity is independent of α, we should expect the formula to hold even when $\alpha = 90°$, i.e., for free vertical fall. And doesn't it? Is not (6) the same equation as (5) but for difference in notation? With the clarity of hindsight we now see that (6), not (5′), is the truly enlightening analogue. The questions of remarkable men have remarkable answers: even Galileo was astonished.

3.1.5 Conservation of Energy

From (6),

$$\tfrac{1}{2}V^2 = gH,$$

and introducing mass m, the substance whose attraction by the Earth's gravitational field (i.e., whose weight) is mg, we have

$$\tfrac{1}{2}mV^2 = mg \cdot H. \tag{7}$$

And what is $\frac{1}{2}mV^2$? Yes, the kinetic energy, the energy of the motion. And $mg \cdot H$? mg is the force exerted by gravity on the substance m, so that $mg \cdot H$ is the work done against gravity in raising mass m a height H. When so raised, although not in motion, m has capacity for motion; it has, as we say, potential energy. When m is let fall, its stored energy is utilized to produce motion; what was potential becomes kinetic. The loss of the former is the gain of the latter. There is no overall loss, the total of used and ready-to-be-used energy remains unchanged; the energy is conserved.

Although Galileo came close to formulating this concept, it nevertheless escaped him—and his successors for more than two centuries. He fully appreciated the implications of (6), but not those of (7). He did know that the motion of (6) is reversible; that if a body in sliding down a perfectly smooth plane from rest loses height H in reaching the bottom with velocity V, it will when projected from the bottom with velocity V just reach the top of an incline of height H. This he demonstrated by letting a body slide down one incline and up another of the same height. See Fig. 3.8.

To prevent the body sliding down one incline from jamming against the edge of the other at A, it is of course necessary to round off the corner at A. Actual conditions being less than ideal, there being some friction despite smoothed and polished inclines, the particle from O does not

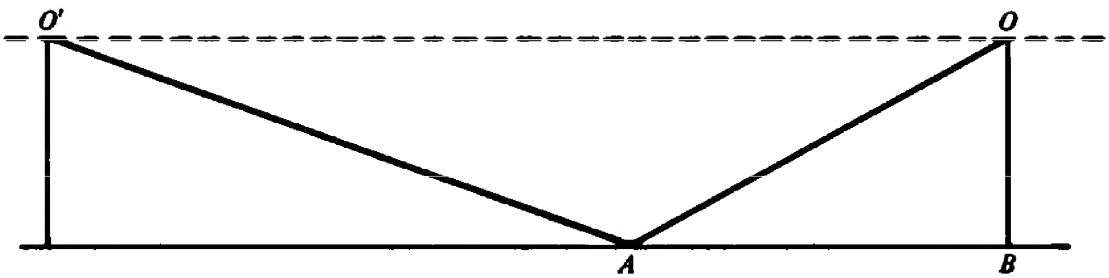

Figure 3.8

quite succeed in reaching O'. Were it successful it would return to O, and from O to O' again—and again. Perpetual motion is an idealization, not a reality. A fact that reminds us that absence of friction is essential to conservation of energy. With friction, some potential energy is changed not into kinetic energy, but into heat.

To eliminate friction Galileo made an experiment justly regarded as a classic. What is needed—apart from genius—to conceive it? Two nails, string, a heavy bob, and a lighted candle. See Fig. 3.9. Why the candle?

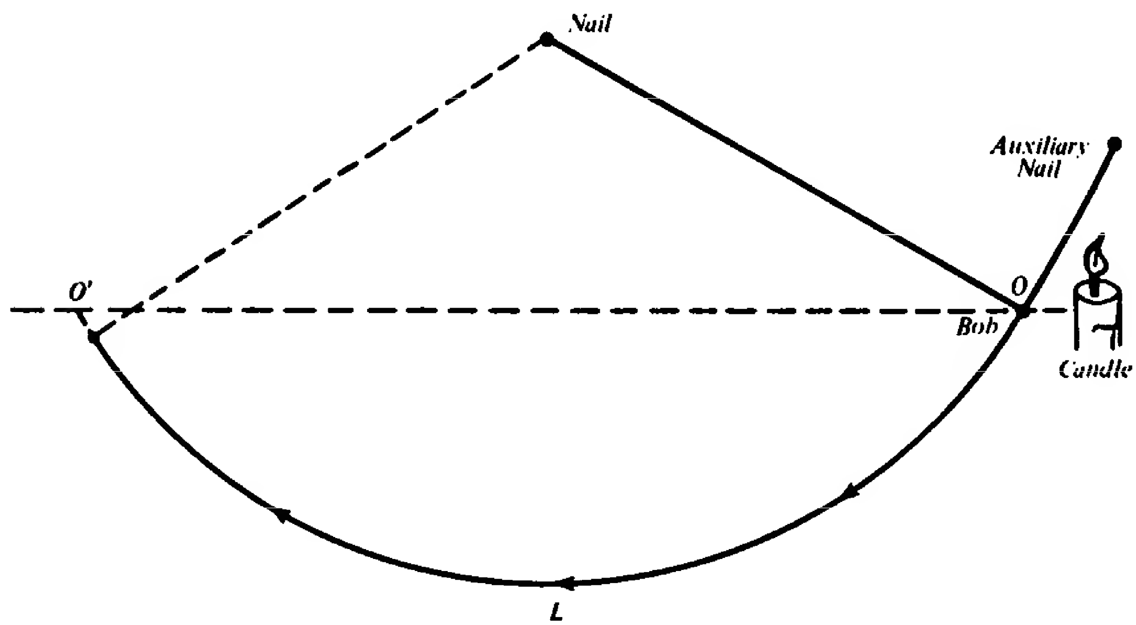

Figure 3.9

To burn through the auxiliary string, thereby releasing the pendulum from rest. Although ever so careful in releasing by hand, one might inadvertently pull back or push forward; the object is to let the bob start of its own accord. What happens? The bob swings down from O to L and then back up—almost—to the same level at O'. Almost, but not quite to the same level, because there is just a little friction between the not perfectly flexible string and supporting nail and friction of air resistance to motion of both string and bob. In thus avoiding the relatively gross

friction of a pair of inclined planes, Galileo's experiment closely approximates the ideal.

Did you anticipate the result? Oh yes, I know you are tediously familiar with the swinging of a pendulum. The point is: did you anticipate this result *as a consequence* of the inclined plane result (illustrated by Fig. 3.8)? Or did you from the all-too-familiar swing of the pendulum infer the result for (idealized) inclined planes? It takes genius to see the commonplace with discerning eyes.

What did Galileo see? At any point P of its circular arc the bob is moving, momentarily, tangentially to the circle at P. It is, in effect, moving—just for a moment—along a very short segment of an inclined plane whose slope is that of the tangent at P. For other moments the bob is moving along other inclines; other inclines with other angles of inclination. But, precisely because the motion is independent of the angle of inclination it matters not whether the bob traverses two or two hundred planes. Isn't the circular path a limiting case where the motion takes place along infinitely many planes? Fig. 3.10 is suggestive. Think about it. Have you ever discerned the swinging of a pendulum as motion along infinitely many inclined planes? More important, would you have seen its implications?

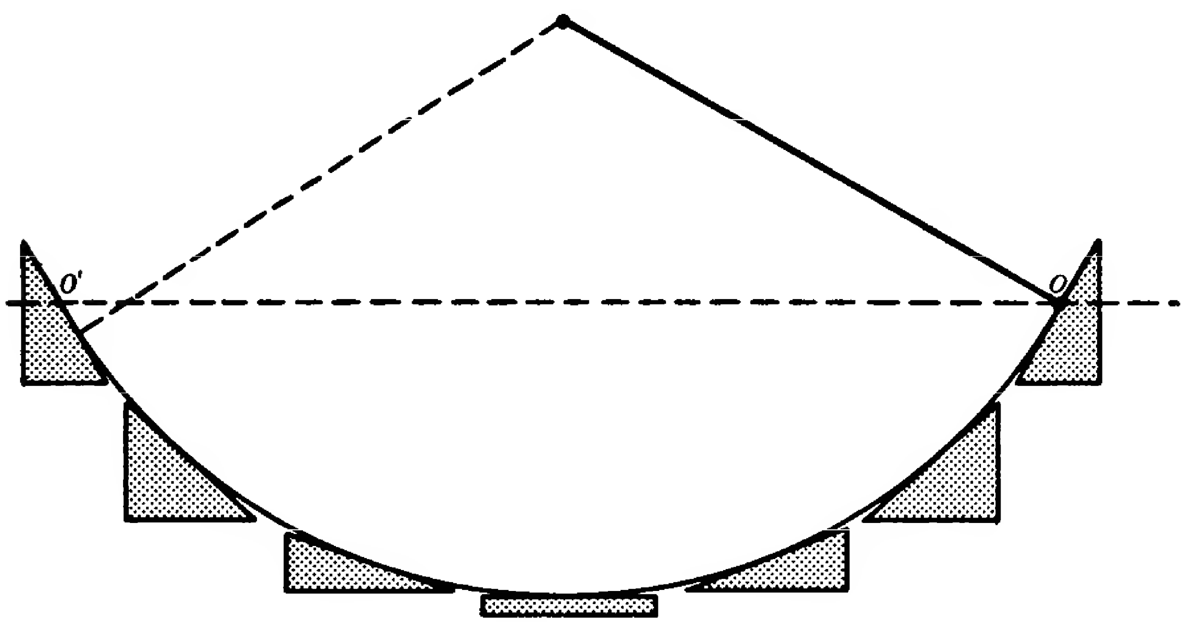

Figure 3.10

Vary the data. Galileo did. To repeat his additional series of experiments we need additional apparatus; we need another nail. See Fig. 3.11. The extra nail N_1 is fixed vertically below that suspending the bob.

What happens? When the bob reaches the lowest point L its suspending string gets caught against N_1, so that the subsequent motion of the

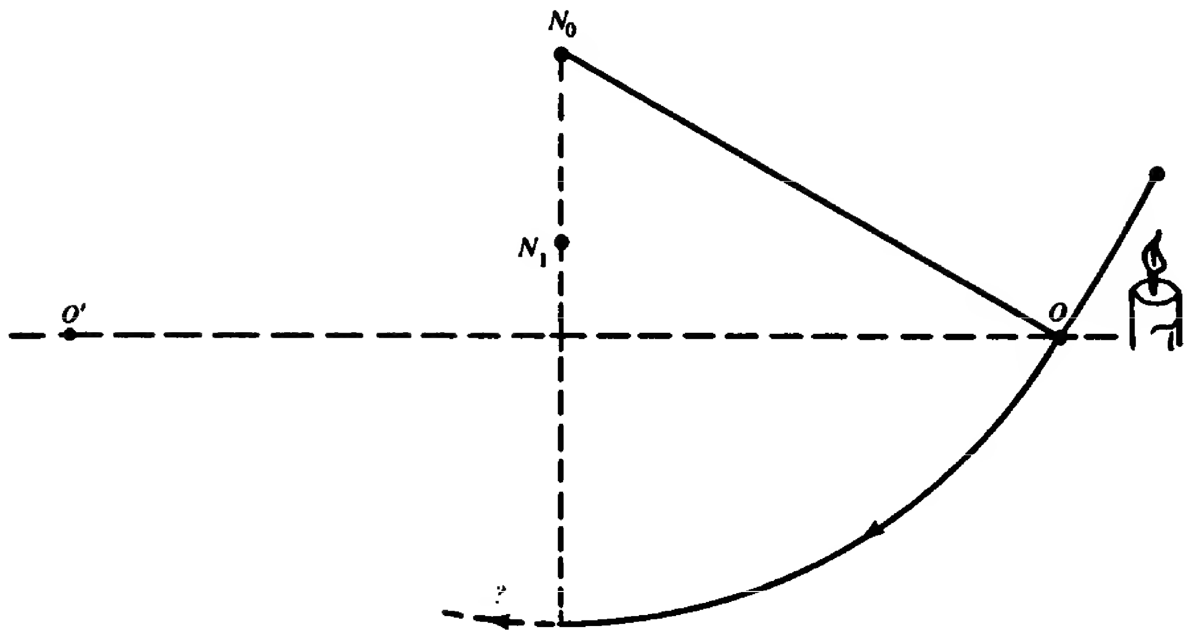

Figure 3.11

bob is along a circular arc of radius N_1L about N_1 instead of N_0L about N_0. But its motion at L is simultaneously tangential to both circles since here they have a common tangent, so that there is no disruption of the continuity of its motion. There being no disruption, there is no loss of velocity. There being no loss of velocity we expect the bob to ascend to *almost* its original level. (Remember air resistance and the imperfect flexibility of the string.) It does. Additional confirmation is reassuring. See. Fig. 3.12.

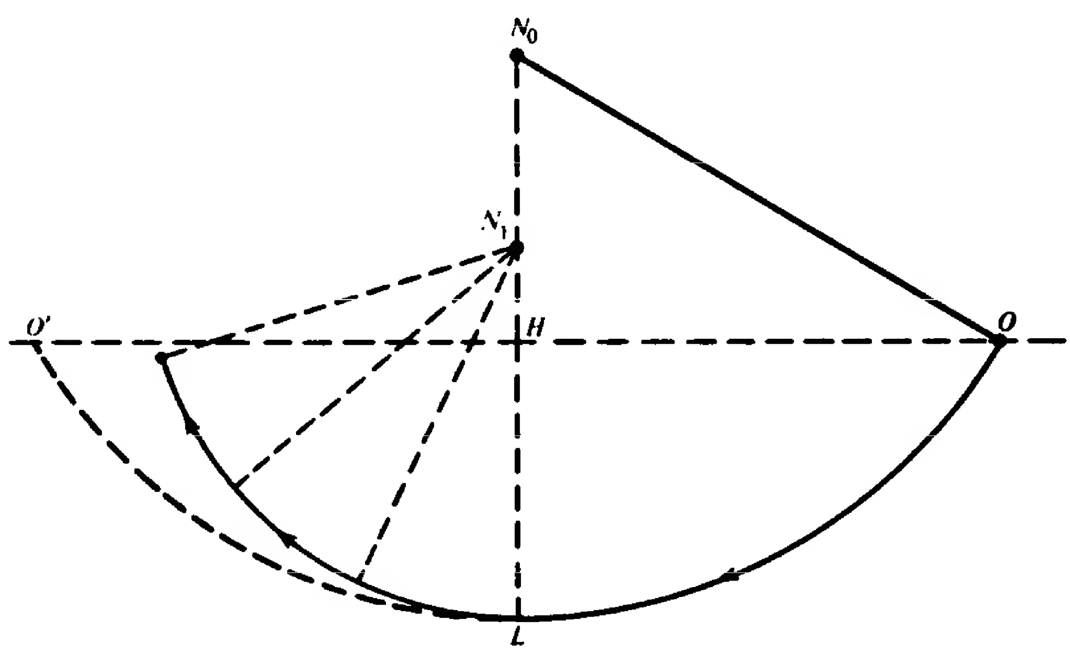

Figure 3.12

Continue to vary the data. Vary the height of N_1 above L. What happens when N_1 is midway between L and H? The bob just about reaches H. It has, in effect, climbed planes of all angles between 0° and 180°. And if N_1 is nearer to L than to H? See Fig. 3.13.

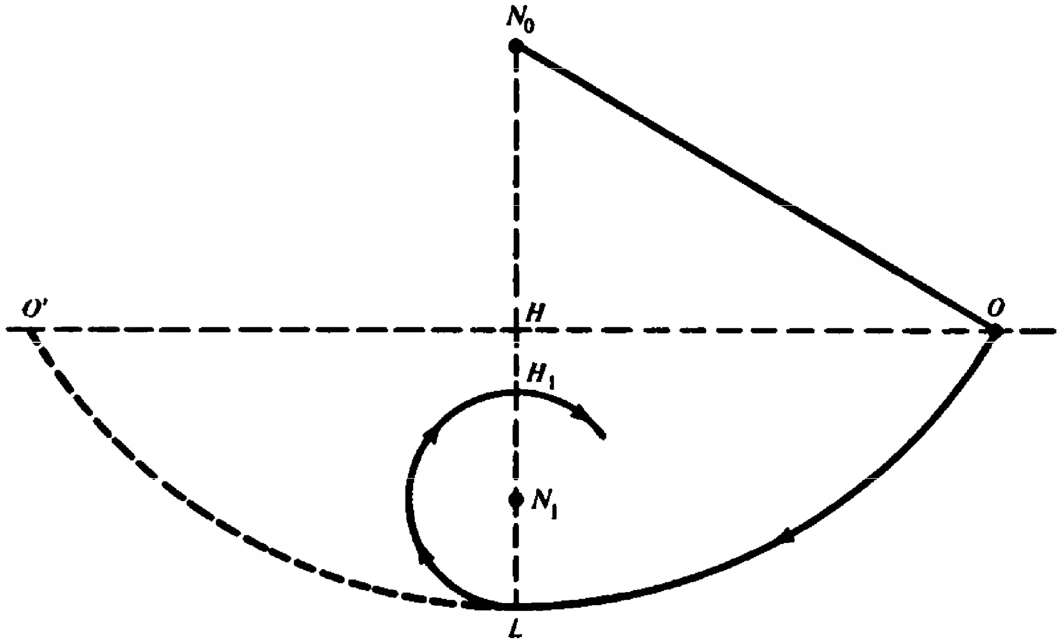

Figure 3.13

The physical constraint of the string prevents the bob from going any higher than H_1, where $H_1N_1 = N_1L$. The bob is carried beyond H_1 and the string begins to wind around the nail N_1, thereby demonstrating that the bob has a residual velocity at H_1, that it would have gone higher but for the constraint.

Here is verification, elegant in its simplicity, that a falling body acquires sufficient velocity to return to its original height. Do not be deterred from making these experiments for want of a nail; for want of a nail a kingdom was lost. The whole apparatus can be purchased for a quarter of a dollar or half a crown. Yet remember it was the man behind the experiment who made what could be an idiot's plaything one of the great experiments of physics.

3.1.6 Law of Inertia

What more is there to say? That depends upon whether you think about these experiments with Galileo's intelligence. We reconsider the situation illustrated by Fig. 3.8. We know that with idealized planes a body let slide at O would regain its original height no matter what the inclination of AO' to the horizontal. Now suppose AO' to be nearly horizontal. What happens? The slope is so gentle that in regaining its

height the body has to travel miles and miles up the incline. The more nearly the incline is to dead level, the farther along it the body will slide to regain its original height. If the incline is precisely horizontal, the body will have to travel on, and on, and on.

What about its velocity? We all know, willy-nilly, without experiment, that the steeper the incline, the greater the deceleration of the body ascending it; the gentler the incline, the more slowly a body ascending it will lose speed. If AO' is only just uphill the dropping off of speed must be a very gradual affair, yet if AO' were downhill ever so slightly there would be an increase of speed. So, what happens if AO' is dead level? There can be neither a slowing down nor a speeding up. So? The body must continue at constant velocity. And how far along the horizontal incline must it go to regain the height of O? It has to go on, and on, and on. So? It must go on, and on, forever, and ever.

Galileo's theory being consistent with our common experience, we anticipate that these conclusions may be drawn from his theoretical equations as well as from his experiments. From (1′) in Number 3.1.4, we obtain the equation

$$\frac{V}{g \sin \alpha} = T;$$

in other words, a body with velocity V at the foot of a plane of inclination α will, under idealized conditions, reach the top of the plane in time T. Consequently, since $\sin \alpha$ tends to 0 as α tends to zero, T becomes infinitely large when the incline becomes horizontal. Also, we recall that the body in question is, in effect, falling freely in a gravitational field $g \sin \alpha$, i.e., with an acceleration $g \sin \alpha$. When $\alpha = 0$, $\sin \alpha = 0$, so that $g \sin \alpha = 0$, i.e., there is no change of velocity.

What from all these experimental and theoretical considerations do you conclude? Galileo's conclusion is the Law of Inertia. A body will continue in its state of rest or of uniform motion in a straight line until acted upon by external forces (e.g., gravity, friction) to change that state.

The astute reader may protest that we have tacitly used the Law of Inertia in deducing it. Such protest misunderstands the situation: Galileo was not making deductions from established theory; he was establishing a theory. Tacit use is a step towards explicit use; inarticulate experience a step towards articulated experience. And the steps? Varying the data in accordance with the concepts of a fertile imagination.

But why is this law described as *Law of Inertia*? An inanimate body, unlike a person or an animal, does nothing to control its own motion. Whither it goes and how it goes are at the mercy of external forces. It is inert.

Galileo invariably considered the Law of Inertia within the context of his discovery; he always thought of uniform motion along a straight line in an infinite plane. Never could he escape the terrestial; his thoughts were Earth-bound. Of course he knew that the Earth is spherical, yet he never thought out the consequences. He knew what little in his time there was to know about the stars, he was one of the first to use a telescope, but he never came to the idea of applying his Law of Inertia to the stars. A simple idea, yet a tremendous jump forward. It is as if Galileo became a victim of his own law, bound by the inertia of a fixed context. It is remarkable Galileo did not make the jump; it would have been more remarkable had he done so. Galileo was Galileo, not Newton.

3.1.7 A Cannon Ball's Trajectory

It was in Galileo's time that firearms were invented; cannon became the final argument of kings. Although a deadly subject, the efficacy of new methods of killing for one's country is always a lively issue. What is the path of a cannon ball? The question was of great scientific interest as well as of practical importance. Characteristically, Galileo was engrossed by the problem; characteristically, he solved it. The outcome of his ingenuity we know today as the method of superposition.

Like Galileo, to reduce the complexity to the manageable, we neglect the dimensions of the cannon ball and consider it to be merely a material point. To simplify further we neglect friction although air resistance to a cannon ball is by no means negligible. Galileo did not have the means for precise measurement, and remember that a first approximation is a step towards a better approximation. Unlike Galileo, we are able to facilitate his solution by using a little algebra and an orthogonal coordinate system. It is vital to his solution that the one axis is horizontal and the other vertical. See Fig. 3.14.

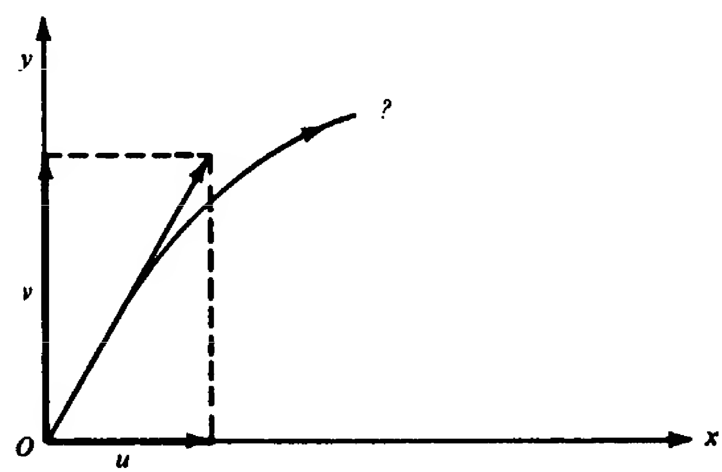

Figure 3.14

The initial velocity of the cannon ball when leaving the cannon's mouth is represented in magnitude and direction by the big vector from 0. This vector is resolved into a component u along the horizontal x-axis and a component v along the vertical y-axis. (Note that the letter u comes before v as does x before y, so that u is associated with x and v with y. Respect the alphabet!) So, timing the ball's motion from the cannon's mouth, at $x = 0$, $y = 0$, when $t = 0$, its horizontal velocity is u and its vertical velocity v. What are its component velocities in these directions at time t?

Galileo's deep insight is that the horizontal motion is unchanged. The horizontal component of the ensuing motion is that of a particle traveling in a gravitationless field. Remember his Law of Inertia. This component remains u. So, at end of time t, the horizontal displacement x is given by

$$x = ut. \tag{8}$$

And what about the vertical component of the motion? Likewise, if there were no gravitational pull vertically downwards, we would have

$$y = vt.$$

But this is contrary to fact, so let us be mindful by writing the letter y with a subscript 1, viz.

$$y_1 = vt. \tag{9'}$$

Next, taking gravitation into account and ignoring the initial velocity, from (4) we have

$$y = \tfrac{1}{2}gt^2$$

where the positive y-axis is vertically downwards. So, with positive axis vertically upwards,

$$y = -\tfrac{1}{2}gt^2.$$

Yet, to be mindful of our neglect of the initial velocity v, we write the letter y with another subscript,

$$y_2 = -\tfrac{1}{2}gt^2. \tag{9''}$$

It is at this stage that Galileo makes use of the principle of superposition. He argues that the total upward displacement y (in time t) of a particle leaving O with initial velocity v and decelerated by gravity will

be the sum of the displacements y_1 and y_2, i.e., of the displacement (in time t) with initial velocity v but no gravitational field and the displacement (in time t) with gravitational field but no initial velocity. What is his argument? First, suppose the displacements to take place consecutively; in time t the particle is displaced y_1; subsequently in a similar time the particle is displaced a farther distance y_2. Clearly, the resultant of the consecutive displacements is their sum, the joining on or adding in position of the latter to the former. In short, superposition is obviously applicable to the displacements resulting from the successive motions. The crux of the matter: Is superposition applicable to the resultant displacements if the motions occur simultaneously? Whether or not a particle has an initial velocity is independent of the presence or absence of a gravitational field, and a gravitational field is independent of whether or not a particle has an initial velocity. So surely both motions may occur simultaneously without either altering the other, so that the displacements due to these motions are unchanged by the simultaneity of the motions. Superposition is still applicable; from (9′) and (9″) we have

$$y = vt - \tfrac{1}{2}gt^2. \tag{9}$$

We can describe completely the cannon ball's trajectory if we can always answer the question: Where is the cannon ball now, t seconds after being fired? The pair of equations (8), (9) give precisely this answer; from (8) we get its present horizontal displacement x, from (9) its present vertical displacement y; i.e., we get its present position (x, y). Told when, we can compute where. For any "when" (t) we can plot the "where" (x, y) and so obtain a picture of the cannon ball's path.

If we know the cannon ball's present horizontal displacement x, by (8) we can find when it was fired and so by (9) find its present vertical displacement; given x we can compute the corresponding y via the go-between t. Mathematically speaking, t is said to be a parameter, x and y are said to be given parametrically. Somewhat analogously, if X is the father of T and Y is the only son of T, T is the parameter, the middleman between Y and X. Eliminating reference to T, the middleman, we have that Y is a grandson of X. It would be convenient to have y deal directly with x. Can we get rid of the middleman t? From (8)

$$t = \frac{x}{u},$$

so that

$$t^2 = \frac{x^2}{u^2}.$$

Substituting for t and t^2 in (9), we get

$$y = v \cdot \frac{x}{u} - \frac{1}{2} g \cdot \frac{x^2}{u^2}. \tag{10}$$

What sort of curve is given by (10)? Can we transform this equation into a more familiar pattern where the graph is known? To make the coefficient of x^2 equal to 1, we multiply through by $-2u^2/g$, obtaining

$$-\frac{2u^2}{g} y = -\frac{2uv}{g} x + x^2.$$

Adding u^2v^2/g^2, the square of half the coefficient of x, to each side, we get

$$\frac{u^2v^2}{g^2} - \frac{2u^2}{g} y = \frac{u^2v^2}{g^2} - \frac{2uv}{g} x + x^2,$$

i.e. (the square completed)

$$-\frac{2u^2}{g}\left(y - \frac{v^2}{2g}\right) = \left(x - \frac{uv}{g}\right)^2.$$

But this is of the form

$$-\frac{2u^2}{g} Y = X^2,$$

where

$$Y = y - \frac{v^2}{2g} \quad \text{and} \quad X = x - \frac{uv}{g},$$

i.e., a parabola with vertex $X = 0, Y = 0$, and axis $X = 0$. When $X = 0, x = u \cdot v/g$, and when $Y = 0$, $y = v^2/2g$, so that (10) is the equation of a parabola with vertex $(uv/g, v^2/2g)$ and axis $x = uv/g$.

At first sight the coordinates of the vertex seem unenlightening. What is their physical significance? Think back. By equation (5) we know that a particle projected vertically with velocity v will just reach a height of $v^2/2g$; the vertex is at the maximum height of the trajectory. But a parabola is symmetrical with respect to its axis, so? Why, we must expect uv/g to be half the cannon's range. Is it? When the cannon ball returns

to the horizontal plane, $y = 0$; i.e., in (10)

$$0 = \frac{x}{u}\left(v - \frac{g}{2} \cdot \frac{x}{u}\right).$$

But (the ball having left the cannon's mouth) $x \neq 0$; therefore

$$0 = v - \frac{g}{2} \cdot \frac{x}{u},$$

so that

$$x = \frac{2uv}{g}.$$

Thus uv/g is indeed half the cannon's range. We are now able to complete Fig. 3.14. It becomes Fig. 3.15.

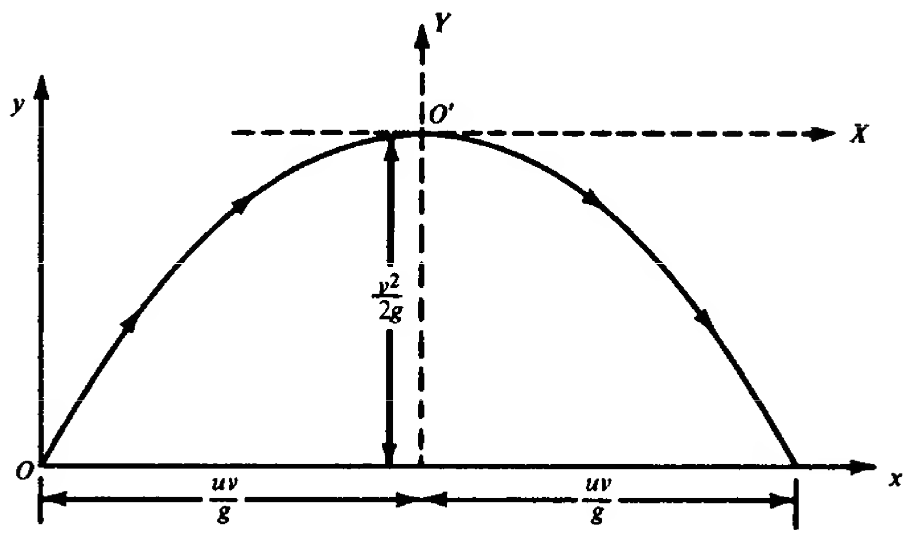

Figure 3.15

How did Galileo verify experimentally that the cannon ball's trajectory is a parabola? Have you not seen performing dogs, and even seals, jump through hoops? Success is assured by placing the hoop where the jumper is going to jump. Galileo used this principle. See Fig. 3.16. The inclined plane is a device to give the ball a predetermined velocity u along the horizontal $A'O'$, so that it hurtles horizontally into space as if at O' in Fig. 3.15. Its neat passage through a series of hoops whose centers are on a parabolic arc confirms his theory.

Yes, a little naive by modern standards, but who with the technology of Galileo's day could devise a better? Speaking of ingenuity, refer back to Fig. 3.5. I never told you how Galileo measured time; watches were

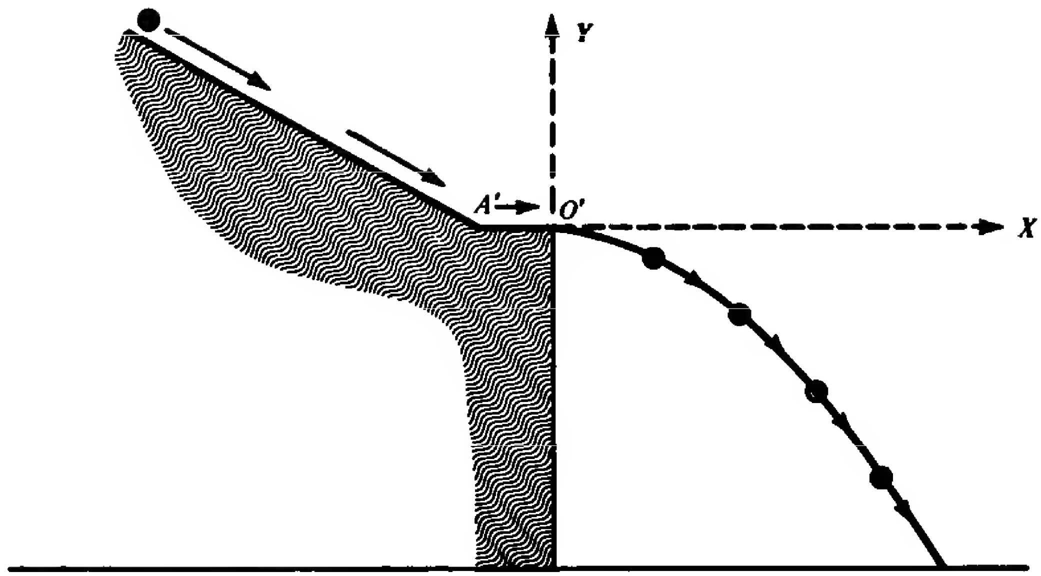

Figure 3.16

nonexistent. He glued tiny slats across the incline at A, B, C, D, big enough for the sliding body to be audible when hitting them, yet not sufficient an obstacle to impede the motion appreciably. His ear his metronome—like his father he was a good musician—he judged the intervals equal. All physicists use their heads; the best also think with their fingers.

SECTION 2. NEWTON

Inevitably, Galileo leads to Newton. Newton was born on Christmas Day, 1642, some eleven months after the death of Galileo; a fact the transmigrationists among my readers cannot fail to remember. Never has Santa Claus brought the world a more enlightening Christmas present. Newton died in 1727, yet the important date for us is 1687. This is the year in which he was finally goaded by his staunch friend Halley into publication of *Principia Mathematica*—Halley paid for the printing. Never before or since in the history of science has a man with so much to say been so reluctant to publish. Leibniz said that of all the mathematics that had ever been done, Newton had done the greater part. This remark was made before they quarreled.

Newton's personality was less colorful and his career less dramatic than Galileo's. Unlike Galileo he was shy and retiring and hated controversy. It is said that when asked to allow his name to be put forward for election to the Royal Society, he at first declined on the grounds that election would necessarily enlarge the circle of his acquaintances. His life is his works.

His body, like Galileo's, lies in the Westminster Abbey of his country; his System of the World, his Law of Universal Gravitation, his mechanics, have become an integral part of educated common sense.

3.2.1 Apples, Cannon Balls, and the Moon

g is for gravity that made the apple in Newton's orchard fall. That this is an old story is certain, that it is a true story is not certain. Certainly it is a good story.

When Newton was a young man up at Cambridge there was a plague. To escape it he retired to his parents' farm at Woolsthorpe in Lincolnshire. There in a year or so of countryside peace he made his greatest discoveries: the concepts of universal gravitation and the infinitesimal calculus. Whether or not he was hit by a falling apple when meditating in the Woolsthorpe orchard, he was certainly struck by a great idea. Although Newton approached the problem of gravitation with an open mind, he did not approach it with an empty mind: thousands of people have seen apples fall without being struck by Newton's idea.

What did Newton have in mind when meditating in his orchard? A diagram in an appendix to *Principia* entitled "The System of the World" must make his train of thought an open secret. He knew certain things about apples, cannon balls, and the Moon, things that were common knowledge to the physicists of his time. The Moon, like the apple, is roughly spherical and presumably heavy, so why doesn't the Moon fall too? The apple is pulled to the Earth by the Earth's gravitational attraction. Why not the Moon? What makes the Moon orbit about the Earth? Galileo's Law of Inertia implies that the Moon would continue with uniform speed in a straight line were it not acted upon by a force to change this motion. What pulls it from its would-be straight-line path to move on a curve concave towards the Earth?

But how on earth can one relate the path of a falling apple to the elliptic orbit of the Moon? Straight lines are so different from ellipses; apple paths so different from Moon paths. Could two curves be more dissimilar? How could both possibly be exemplifications of one law?

Newton saw the possibility; he had the insight of genius. His great idea? Cannon balls. Yes, cannon balls. Had not Galileo shown the trajectory of a cannot ball to be a parabola? Isn't a falling apple a little cannon ball fired with negligible horizontal velocity? So, isn't its trajectory a limiting case of parabolic motion? And the Moon? Isn't this a large cannon ball? Isn't this a large cannon ball fired with great horizontal velocity?

Consider a cannon ball fired toward the Pacific from a mountain peak in the Andes. Given a high muzzle velocity, isn't it conceivable that the ball could be fired right out into the Pacific? With a higher peak to fire

from and a greater muzzle velocity to fire with, wouldn't its trajectory be a larger parabola? Couldn't the ball be fired clear across the Pacific Ocean? But if its trajectory could reach halfway around the Earth, why not three-quarters? Imagination costs nothing; if three-quarters, why not four-quarters? How exciting to see the cannon ball score a direct hit on the cannon from which it was fired. For more excitement, more muzzle velocity. What now? The cannon ball does not land on its cannon after circling the Earth; it blows the gunner's head off and keeps on going. An ending parabolic trajectory is replaced by an unending closed curve; we have a cannon ball moon in orbit.

Newton had the fertility of mind to see the continuous transition from apple to Moon. Surely this must be the most spectacular argument by analogy in the history of science. You will find a copy of Newton's *Principia* diagram in my *Mathematics and Plausible Reasoning*, Vol. I, p. 27.

If the Moon is kept in orbit by a force exerted by the Earth, are not Earth and the other planets kept in orbit about the Sun by a force exerted by the Sun? To let imagination run riot is one thing; to back up highly speculative conjecture by what finally becomes an overwhelming accumulation of supporting considerations is entirely another matter. Newton had the capacity of mind to do both.

3.2.2 Never Smoke Without Fire

Apples and stones fall; the farther they fall, the faster they fall. What causes them to speed up? Supposedly a force exerted on them by the Earth. Yet a force is not something that can be seen. If it cannot be seen how can it be measured? By its effect: smoke is evidence of the valley fire the other side of the hill. What is the effect of a force? Acceleration, the increase in velocity it causes.

From Galileo, for a body falling from rest, we have

$$v = g \cdot t.$$

If time τ later the velocity has increased by β, then

$$v + \beta = g(t + \tau).$$

Subtracting the former from the latter, we have

$$\beta = g \cdot \tau,$$

and

$$\frac{\beta}{\tau} = g;$$

i.e.,

$$g = \frac{\text{increment of velocity}}{\text{increment of time}}.$$

But if, for example, an increment in velocity of 6ft/sec occurs in 3 seconds, this is at the same uniform rate as an increment of 2 ft/sec occurring in 1 second, i.e.,

$$g = \text{increment of velocity per unit time} = \text{acceleration.}$$

We know that Galileo found g to be a constant, yet remembering the necessary imperfection of measurement we must be cautious. At or near the Earth's surface g is a constant within the errors of measurement. It turns out that this answer is a very close, but only a very close, approximation to the truth.

However, the crux of the matter is that acceleration is a measure of force. What bearing has this on the motion of the Earth and other planets around the Sun? What would be evidence that each is kept in orbit by a force exerted on it by the Sun? Its acceleration towards the Sun.

3.2.3 That the Planets do Accelerate Towards the Sun

Let us suppose that the Moon accelerates towards the center of the Earth and that the planets accelerate towards the center of the Sun. What are the consequences of these suppositions? What sort of orbit will Moon and planet have? This is a hard mathematical question because the acceleration takes place continually and is therefore difficult to take into account. How are we to deal with continual acceleration? Well, how did Galileo deal with continually increasing velocity? Look at Figs. 3.2, 3.3, and 3.4 again and think about them.

Yes, Newton as Galileo, and we as both Newton and Galileo, must deal with the continual, the continuous, the gradually changing by starting with a caricature, discontinual, discontinuous, discrete jerky jumpy change and then by increasing the number and decreasing the jerkiness of the jumps, make the jerky change become less and less perceptibly different from gradual change. Thus fiction becomes reality. To treat the continuous as limiting case of the discrete is really the fundamental idea behind the integral calculus. Newton invented it precisely to facilitate this treatment. Of course, he inherited much from Archimedes, from Cavalieri and from Fermat, yet his contribution was definitive. History justly claims him as a founder of the calculus.

We may be certain that Newton obtained his results in mechanics by integral calculus, but in his published exposition, *Principia Mathematica*,

he insists on not using calculus—even though he invented it for himself. He argued that his readers would find his mechanics shock enough without the difficulty of learning calculus. Whether this made *Principia* easier reading for his contemporaries I cannot tell you; unquestionably it makes it harder for us. The elementary but obsolete methods used therein can compete with the calculus as successfully as the abacus with the electronic computer. However, fortunately for both author and reader, *Principia* contains one deduction which is as simple as it is important. It is the answer to the question posed above: What is the orbit of a planet which continually accelerates towards a fixed point C? We turn to Newton's answer.

We suppose a planet P the instant it is at A_1 to be moving at v ft/sec. Furthermore we suppose that during the ensuing second it is not acted upon by any external force. What happens? In accordance with Galileo's Law of Inertia it continues to move uniformly in a straight line with a velocity of v ft/sec. Consequently, since velocity is space traversed in unit time, one second later it is at a point A_2, v feet from A_1, and the directed line segment A_1A_2 represents in both magnitude and direction its velocity during this second.

Also suppose that the moment P reaches A_2 it receives an instantaneous acceleration towards C (C is for *Center*, say, the center of the Sun). What does not happen? Had our planet not received this acceleration when at A_2, in accordance with the Law of Inertia, it would, of course, have continued at v ft/sec along A_1A_2 (produced) forever. One second later it would have been v feet from A_2, at F_3 such that $A_1A_2 = A_2F_3$ (F is for *Fictitious*—P never actually gets to F_3). See Fig. 3.17. But when at A_2 our planet receives an acceleration towards C which is represented in magnitude and direction by (say) the directed line segment A_2C_3. Thus the resultant velocity of P on leaving A_2 is represented by the diagonal A_2A_3 of the vector parallelogram illustrated by Fig. 3.18.

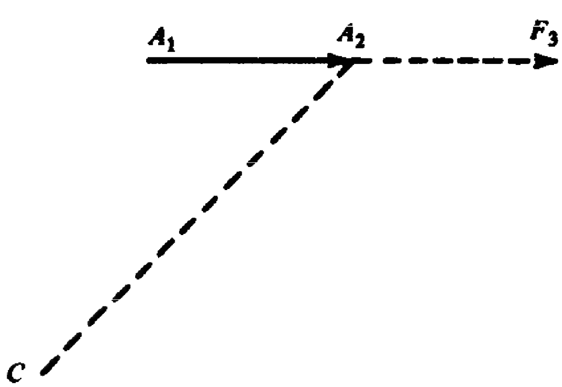

Figure 3.17

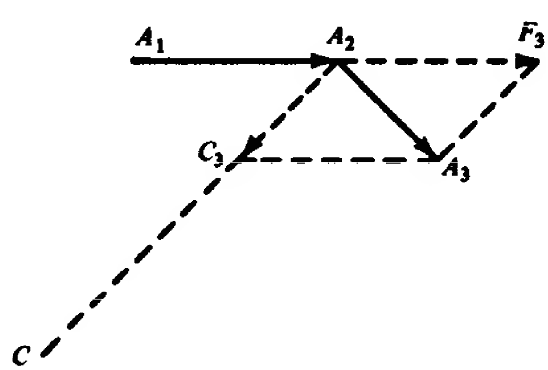

Figure 3.18

What happens afterwards? Remember that the accelerating impulse which acted on P when at A_2 acted only for an instant. When P left A_2 this impulse no longer acted. So, after leaving A_2, in accordance with the Law of Inertia, P continues to move in the direction A_2A_3, traversing a distance A_2A_3 every second—until it is subjected to another external impulse. One second after leaving A_2 our planet actually reaches A_3 (A is for *Actual*).

At A_3 we suppose our planet to receive another instantaneous impulse towards C. Similarly, this causes a second instantaneous change of velocity. The actual velocity of P on leaving A_3 is likewise represented by the diagonal A_3A_4 of the vector parallelogram illustrated in Fig. 3.19.

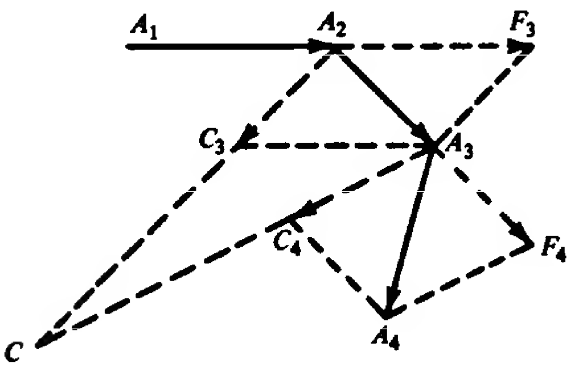

Figure 3.19

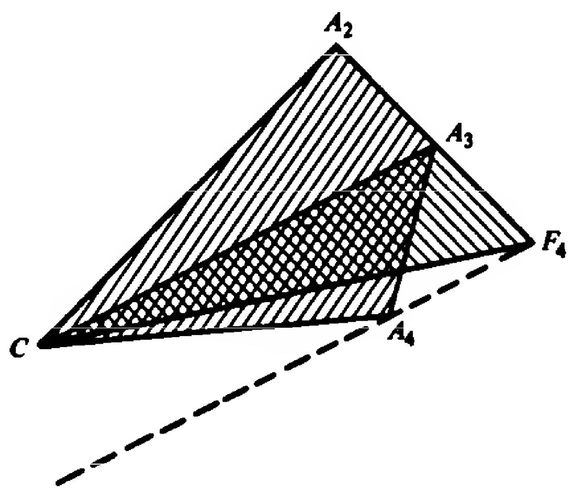

Figure 3.20

Had P not received an accelerating impulse towards C when at A_3, it would by the Law of Inertia have continued along A_2A_3 (produced) to reach F_4 a second later. Had P not had a velocity $\overrightarrow{A_2A_3}$ when it received at A_3 an acceleration towards C, it would have moved along A_3C to reach C_4 a second later. With both velocities P actually travels along A_3A_4 and reaches A_4 a second later. It is unnecessary for us to consider the next accelerating impulse given to P at A_4.

Careful consideration of the consequences of these discrete discontinual accelerations of P towards C is the key to determination of the consequences of continual acceleration of P towards C.

Note that in Fig. 3.19, $A_2A_3 = A_3F_4$ and that $F_4A_4 \| A_3C$ (because $\|$ to A_3C_4). Let us redraw this figure with a new emphasis: Fig. 3.20. Because $\triangle$'s CA_2A_3, CA_3F_4 have equal bases A_2A_3, A_3F_4 and the same altitude, they are equal in area. Symbolically,

$$\triangle CA_2A_3 = \triangle CA_3F_4.$$

And because $\triangle$'s CA_3F_4, CA_3A_4 have the same base CA_3 and equal altitudes (because they lie between the same parallels F_4A_4, A_3C), they also are equal in area:

$$\triangle CA_3F_4 = \triangle CA_3A_4.$$

Therefore,

$$\triangle CA_2A_3 = \triangle CA_3A_4.$$

What do we conclude? What is the relevance of this result to the known laws of planetary motion? That in two consecutive seconds (actually the second and third, but this is not important) the radius vector joining the center of attraction C to our planet P sweeps out equal areas. But isn't it clear that the argument would hold if we took some other unit of time instead of a second? Alternatively we could take a tenth of a second, or a hundredth, . . . , or a millionth, or a billionth, or a trillionth, or . . . . As the equal intervals decrease in duration, the effect of jerky jumpy discrete central accelerations differs less and less perceptibly from that of continual central acceleration.

What must we conclude? That if a planet P has a continual central acceleration towards C, then its orbital motion is such that its radius vector PC sweeps out equal areas in equal times. *But this is precisely Kepler's Second Law.*

Seldom has such a simple argument had such important repercussions. It convinced Newton—and should convince you—that the planets are accelerated towards the Sun. Regard Fig. 3.19 and 3.20 with respect: they link together the mechanics of terrestial and interplanetary space.

3.2.4 What is the Law of Universal Gravitation?

We have seen how Newton by discerning a continuous transition between the fall of an apple, the trajectory of a cannon ball, and the orbit of a planet was led to conjecture that the planets have accelerations towards the Sun as do falling apples towards the Earth. And how did he adduce strong support for his conjecture? By showing that Kepler's Second Law is a necessary consequence. (It is just possible that Kepler's Law could be a necessary consequence of an alternative conjecture.)

What is the next step? Granted that the planets do accelerate towards the Sun and the Moon towards the Earth, surely, because of the regularity of their orbits, these accelerations cannot be haphazard affairs, but must be subject to some law. And isn't the whole point of Newton's insight the *continuity* of the transition? Surely similar effects have similar causes. Surely the Earth's gravitational pull on the apple, the cannon ball, and the

Moon must be of the same nature as the Sun's gravitational pull on the Earth. Surely there must be a Law of Universal Gravitation. The next step is to specify it.

At school, I was cheated by my physics teacher. Conjurer Newton puts up the most spectacular show on Earth—or the Solar System—by producing the gravitational rabbit from the universal hat. And what did I get? A bland statement of the Law of Universal Gravitation without any indication of how the trick was done.

How did the rabbit get into Mr. Newton's hat? To appreciate his legerdemain you must first learn a basic trick of orbital conjuring, namely deduction of the central acceleration of a body moving with uniform circular motion. The most elegant method of dealing with the latter was found, subsequent to Newton's derivation, by Sir William Rowan Hamilton (1805–1865) the inventor of the calculus of quaternions. To Sir William's method we now turn.

3.2.5 Uniform Circular Motion: Hamilton's Hodograph

We consider a particle (or planet) P to move with uniform motion in a circle of radius r and center C. Since P moves in a circle, the *direction* of its motion (tangential to the circle) is, of course, continually changing, and consequently, its *velocity* is continually changing also. But, since P's motion is uniform circular motion, it moves equal distances in equal times, that is, its speed v (the distance it goes in unit time irrespective of direction) is constant. Therefore, if the velocity of P at P_1 is represented in both magnitude and direction by $\overrightarrow{P_1A_1}$ and its velocity at any other point P_2 is represented in both magnitude and direction by $\overrightarrow{P_2A_2}$, then both directed line segments must have the same length. Furthermore, this common length must be equal to v, the distance traversed in unit time. See Fig. 3.21.

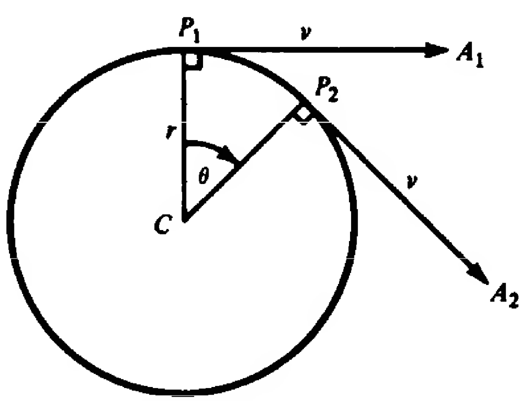

Figure 3.21

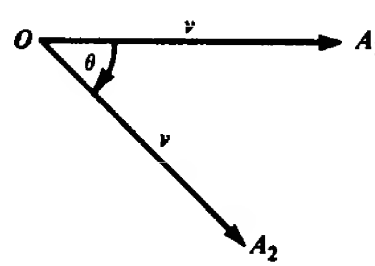

Figure 3.22

That the velocity vectors $\overrightarrow{P_1A_1}$, $\overrightarrow{P_2A_2}$ have the same magnitude was of great interest to Hamilton. What is its significance? Yes, of course, it

implies that P moves with constant speed. But what other significance does it have? Hamilton introduces a new representation to exhibit this other significance. He terms it a hodograph.

Suppose duplicates of the vectors $\overrightarrow{P_1A_1}$, $\overrightarrow{P_2A_2}$ to be moved parallel to the originals so as to originate from some fixed point O. See Fig. 3.22. Thus $\overrightarrow{OA_1}$ is the duplicate of $\overrightarrow{P_1A_1}$ and $\overrightarrow{OA_2}$ is the duplicate of $\overrightarrow{P_2A_2}$. And since the duplicates are parallel to the originals, the angle between the duplicated pair is equal to the angle θ between the original pair. Consequently we may think of Figs. 3.21 and 3.22 as the dials of synchronous watches—synchronous in the sense that as an arm CP rotates uniformly from CP_1 to CP_2 on the original dial, an arm OA rotates from OA_1 to OA_2 on the hodograph dial. It follows that OA will rotate full circle from OA_1 back to OA_1 in the same time T as CP rotates full circle from CP_1 back to CP_1.

What follows? In time T, traveling with uniform speed v, P traverses the circumference of a circle of radius r. So,

$$Tv = 2\pi r.$$

Similarly, in time T, traveling with uniform (because synchronous) speed a (say), A travels the circumference of a circle of radius v. So,

$$Ta = 2\pi v.$$

Hence,

$$\frac{Ta}{Tv} = \frac{2\pi v}{2\pi r}$$

i.e.,

$$\frac{a}{v} = \frac{v}{r}$$

so that

$$a = \frac{v^2}{r}. \tag{11}$$

The simplicity of the mathematics belies the subtlety of its interpretation. What is a? a is, because uniform, the speed of the vector tip A at any instant; the instantaneous speed of A. And what is the instantaneous speed of the vector tip A? The answer to this question is Hamilton's ingenious insight. It is the magnitude of the instantaneous rate of change of velocity of the vector $\overrightarrow{OA}$. But, $\overrightarrow{OA}$ is the duplicated vector representation of P's velocity. In short, a is the magnitude of the instantaneous acceleration of P: Acceleration is the velocity of velocity.

Thus, (11) gives the magnitude of P's acceleration. But what is its direction? Because of Newton's argument that the planets do indeed accelerate towards the Sun, you are no doubt prepared to accept the view that the acceleration of P is towards C. However, it readily follows from Hamilton's hodograph that such is the case, thereby bolstering up our conviction.

The motion of A when at A_1, for example, is instantaneously tangential to the circle through A_1 with center O, i.e., perpendicular (down the page) to OA_1 in Fig. 3.22, and consequently parallel to P_1C in Fig. 3.21. So, the accleration of P when at P_1 is along P_1C. But A_1, P_1 are (corresponding) arbitrary points. In short, we conclude that the acceleration of P is invariably towards the center C of its circle of rotation.

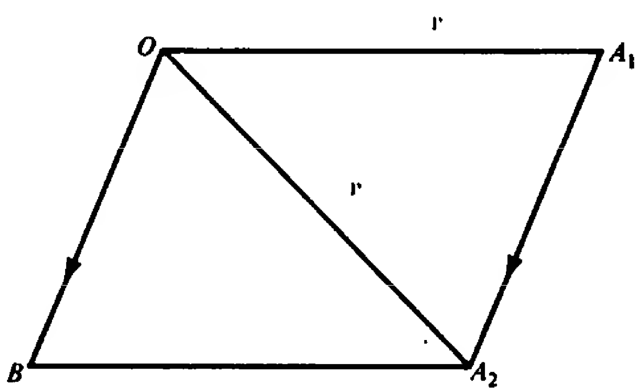

Figure 3.23

What has been said in short about the magnitude of the acceleration may be said at length—at the expense of spoiling a good short story. Suppose that OA is in the position OA_2 time t after being in position OA_1. We complete the vector parallelogram OA_1A_2B. See Fig. 3.23. $\overrightarrow{OA_2}$ is the resultant of $\overrightarrow{OA_1}$ and $\overrightarrow{OB}$, so that the velocity of P at P_1 has to be increased by $\overrightarrow{OB}$ for P to have velocity $\overrightarrow{OA_2}$ at P_2. But this increase of velocity $\overrightarrow{OB}$ occurs in time t, so that $\overrightarrow{OB}/t$ is the average rate of increase of velocity, i.e., the average acceleration of P in moving from P_1 to P_2. But, equally well we may take the equivalent vector $\overrightarrow{A_1A_2}$ instead of $\overrightarrow{OB}$. (Considering the triangle OA_1A_2, velocity $\overrightarrow{OA_1}$ has to be increased by $\overrightarrow{A_1A_2}$ to give $\overrightarrow{OA_2}$). Consequently, $\overrightarrow{A_1A_2}/t$ is the average acceleration of P in moving from P_1 to P_2.

Next, suppose A_2 to be arbitrarily close to A_1. The closer A_2 is to A_1 the more nearly

$$\text{length of } \overrightarrow{A_1A_2} = \text{arc } A_1A_2 \text{ of hodograph}$$

and, consequently, the more nearly

$$\text{length } \frac{\overrightarrow{A_1A_2}}{t} = \frac{\text{arc } A_1A_2}{t}.$$

But,

$$\text{length of arc } A_1A_2 = a \times t$$

so that

$$\frac{\text{length of arc } A_1A_2}{t} = a.$$

Thus, the shorter the interval t, the more nearly are the lengths of arc A_1A_2 and segment A_1A_2 the same; i.e., the more nearly

$$\frac{\text{length } A_1A_2}{t} = a.$$

We conclude that the magnitude of the instantaneous acceleration is a.

3.2.6 On Newton's Discovery of the Law of Universal Gravitation

Newton's great discovery is specification of the relation between the acceleration a of a planet P towards the Sun and its distance r from the Sun, the provision of a formula for a in terms of r. Just this—and the audacity to suppose that every body in the universe exerts an accelerating force on every other body in the universe in accordance with this formula. What is the relevance to Newton's discovery of the formula giving the central acceleration of a body moving with uniform circular motion? A clue is the reminder that an idealization, a good first approximation, often reduces the complexity of a problem to what is manageable. What holds in the simple case may perhaps hold in the general case, or be at least a good indication.

According to Kepler's First Law every planet moves in an elliptic orbit with the Sun at one of the foci. In fact the planets move in elliptic orbits that have very small eccentricity—orbits that are very nearly circles. Mars, of which Kepler made a special study, has a less circular orbit than the other planets except Mercury. It, like the other planets, has minor perturbations or deviations (due to the gravitational attraction of the other planets), yet its orbit is still a very good first approximation to a circle. Introduce simplifying idealization; suppose that it is a circle. What follows?

According to Kepler's Second Law, the motion of a planet is such that its radius vector from the Sun sweeps out equal areas in equals times. But, if the orbit is a circle, then obviously equal areas cannot be swept out in equal times unless the motion is uniform circular motion. So? Why, of course, (11) is applicable. Suppose that R is the radius of Mars' circular orbit about the Sun and v its uniform speed, then by (11) we have that the magnitude a of Mars' acceleration towards the Sun, its centripetal acceleration, is given by

$$a = v^2 \cdot \frac{1}{R}. \tag{12}$$

What follows? We have a in terms of R and v; Newton's problem is to obtain a in terms of R alone. Therefore we must eliminate v; we need a second equation. Cast your mind back for a moment to the derivation of (11). If T is the period of Mars' orbit, then we have a precise analogue to an equation used to obtain (11), namely

$$T \cdot v = 2\pi R,$$

so that

$$v = \frac{2\pi}{T} R.$$

Squaring, and substituting for v^2 in (12), we have

$$a = \left(\frac{2\pi}{T}\right)^2 \cdot R^2 \cdot \frac{1}{R};$$

i.e.,

$$a = \frac{4\pi^2}{T^2} R. \tag{13}$$

We have eliminated v at the expense of introducing T. Are we really any better off? Remember that the existence of a Newton presupposes a Kepler. Hasn't Kepler something, something important, to say about T?

According to Kepler's Third Law, the square of T is proportional to the cube of R. Put alternatively, T is proportional to $R^{3/2}$; algebraically,

$$T = c \cdot R^{3/2},$$

where c is a constant, independent of R. The square of T is

$$T^2 = c^2 \cdot R^3.$$

Substituting for T^2 in (13), we have,

$$a = \frac{4\pi^2}{c^2R^3} \cdot R,$$

i.e.,

$$a = \frac{4\pi^2}{c^2} \cdot \frac{1}{R^2}, \tag{14}$$

so that a is inversely proportional to the square of the distance between Mars and the Sun. This was a great step in Newton's Discovery of the Law of Universal Gravitation.

3.2.7 Scientific Attitude: Verification

The difference between conjecture, hypothesis, theory, and law is a difference of degree rather than a difference of kind. The difference of terminology is one of emphasis, indicative of the well-foundedness of the proposition in question, and consequently, the degree of conviction with which it is held.

The idea that a cannon could, supposing sufficient muzzle velocity, be its own target is a wild conjecture; that a cannon ball could encircle the Earth to return to its own starting point is merely a flight of the imagination. But when such a conjectured flight is seen in the context in which Newton saw it, as an intermediate case between the falling apple's trajectory and the Moon's orbit, its status changes. The aspect of continuous transition gives the conjecture plausibility enough to be considered seriously. What, fancy free, might well have been taken from the pages of *Gulliver's Travels* or *Alice in Wonderland*, might perhaps after all be a physical reality. Wild conjecture becomes sober hypothesis.

When Newton showed that Kepler's Second Law is a consequence of the hypothesis that the planets accelerate towards the Sun, he had a most substantial indication that planets moving with a central acceleration towards the Sun would have the sort of orbits which they do in fact have. Falling apple and orbiting Moon have a common explanation; the terrrestial and planetary pieces of the cosmological jigsaw puzzle fit together. What was entertained precariously is held with some conviction; hypothesis becomes theory. Applying Kepler's Third Law, theory becomes

specific theory: that the centripetal acceleration is inversely proportional to the square of the distance.

Galileo, using the recently invented telescope, discovered that Jupiter has three moons in orbit about it. Later he discovered a fourth. It was found that the period of revolution of Jupiter's moons, as those of the planets around the Sun, satisfy Kepler's Third Law. Here too, planets (i.e. the moons of Jupiter) orbit about their sun (Jupiter) in accordance with the law

$$T = c \cdot R^{3/2}.$$

Here is a second planetary system subject to the same law, the only difference being applicability; each system has its own value of c, the constant of proportionality. These considerations were of great importance to Newton; that Kepler's Third Law also holds is a firm indication of a second planetary system in which centripetal acceleration is inversely proportional to the square of the distance. But if this holds for two planetary systems, why not for a third, a fourth, . . . ? And so Newton was led to his theory of universal gravitation.

But how in physics does theory become law? The act of "legislation" that puts theory on the statute books of physics is verification. And how could Newton make verification? By bringing his theory of the heavenly bodies down to earth, so to speak. Is not this piece of chalk with which I write on the blackboard, as is the Moon, just another planet of the system whose sun is the Earth? But when I let this chalk fall it accelerates towards the Earth's center with terrestial acceleration g. Is the value of g, the central acceleration of our little planet, as deduced from Newton's theory the same as the factual measurement of g? This is the crucial test.

What is the theoretical value of g? Newton deduced it in the following way. By hypothesis, centripetal acceleration of Moon as well as chalk is inversely proportional to the square of its distance from the center of the Earth, i.e., both satisfy the law (equation (14) with notational simplification)

$$\text{centripetal acceleration} = \frac{c}{(\text{distance})^2}, \tag{14'}$$

where c is a constant of proportionality, independent of the distance. (It does the mind no harm to remember that c stands for *constant* and for *centripetal.*) Let R be the distance of the Moon from the center of the Earth and g_M (M is for *Moon*) the Moon's gravitational acceleration

towards the Earth's center; then

(15) $$g_M = \frac{c}{R^2}.$$

Likewise, if r is the radius of the Earth, and consequently the distance of my chalk from the Earth's center, and g_E (E is for *Earth*) my chalk's gravitational acceleration towards the Earth's center,

(15′) $$g_E = \frac{c}{r^2}.$$

Supposing the Moon's orbit about the Earth, as Mars' orbit about the Sun, to be a circle, and consequently by Kepler's Second Law, its motion to be uniform circular motion, say v, and recalling (12) of Hamilton's hodograph,

$$g_M = \frac{v^2}{R},$$

we deduce from (15),

$$\frac{c}{R^2} = \frac{v^2}{R},$$

and thus

$$c = v^2 R.$$

Also, since the Moon's motion is uniform circular motion, its orbital period T is related to v by

$$v = \frac{2\pi R}{T}.$$

Squaring, and substituting for v^2 in the preceding equation, we get

$$c = \frac{(2\pi R)^2}{T^2} \cdot R,$$

that is,

$$c = \frac{4\pi^2 R^3}{T^2}.$$

Unfortunately c cannot be measured directly, but g_E can. From (15′)

$$g_E = \frac{4\pi^2 R^3}{T^2 r^2}. \tag{16}$$

This expresses the gravitational acceleration at the Earth's surface in terms of the quantities r, R, T known to Newton. The principal ingredients of this deduction are given in the diagram shown in Fig. 3.24.

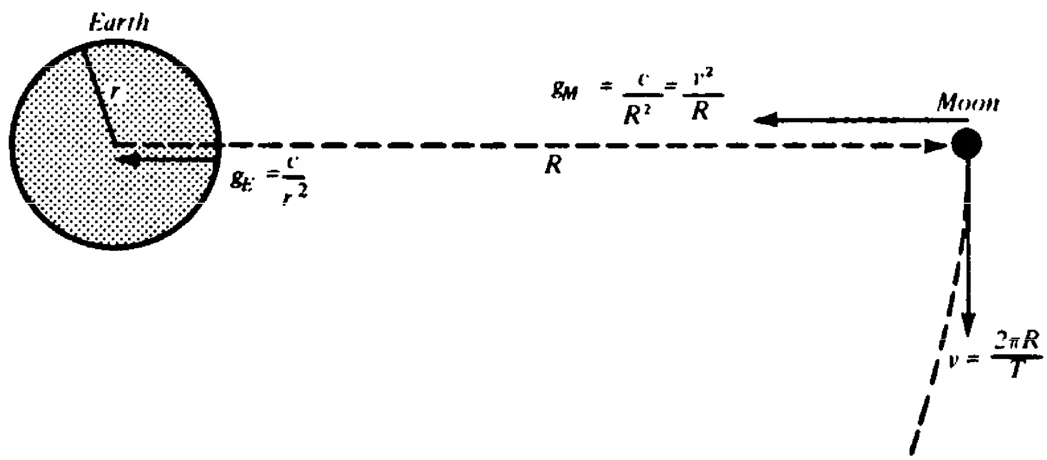

Figure 3.24

Does Newton's theoretical value for g_E coincide with the experimental value? Does the formula check? This you can find out for yourselves. Newton's data are:

the radius of the Earth $= r = 6.3784 \times 10^6$ meters

the distance of the Moon from the Earth $= R = 384.4 \times 10^6$ meters

(so that the Moon's distance from the Earth is about 30 times the Earth's diameter) and

the period of the Moon's orbit $= T = 27.322$ days.

All the data are given to 5 significant figures, except R, which is given to 4. So work to 5 figures and your answer (supposing no arithmetical mistakes) will be reliable to 4. Secondly, if you know the dimensions test, apply it to (16) to check that g_E is the sort of quantity it ought to be, namely an acceleration, LT^{-2}. I shall illustrate the importance of this procedure in the next section, Section 3. One more question: How can g_E be determined experimentally? Yes, by a pendulum experiment. This

also we will consider in the next section. g_E is 9.806 meters/second^2. (Cf. *Principia*, Book III, Proposition IV.)

To his consternation, when Newton did the arithmetic the answer did not come out close enough to the observed value. This set him back eighteen years. The theory must fit the facts; this is the scientific attitude.

We must mention that Newton was reluctant to publish for personal reasons. Sensitive, reserved, indeed somewhat of a secretive nature, he had a strong distaste for controversy—and with good reason. His previous publication of his *Optics* led to a violent quarrel with Hooke who was all too apt to be as bitter as he was brilliant; and Newton's discovery of the calculus led to similar unpleasantness with its other discoverer, Leibniz. While Newton would have been reluctant to publish his *Principia* for fear of further controversy, the discrepancy between his derived value of g_E and the factual value was in itself sufficient reason for him not to publish. Because wrong in an important particular, he would not publish; yet even if right he would have been reluctant.

So much of Newton's theory fitted so well that he asked himself if the data applied to (16) were well determined. T, the period of the Moon, was known with fair accuracy from Babylonian and Greek times; the determination of r and R, considered earlier in these lectures (cf. Eratosthenes), although only roughly determined by the Greeks, were known with but slightly better accuracy in Newton's day. He decided that r was probably ill determined, and awaited its redetermination by a scientific expedition of the French Academy to South America for this purpose. Their evaluation of r gave his theoretical value of g_E close agreement with the experimental; the theory, but not Newton, was ready for publication. Finally, at Halley's insistence and expense, *Philosophiae Naturalis Principia Mathematica* was published. Is there an inverse square law of publication that an author's urge to publish is inversely proportional to the square of his work's merit?

The thoughtful reader will note a few neglected circumstances: for example, formula (16) is derived on the assumption that the Moon moves uniformly in a circle, yet if Newton's theory is correct the Moon's motion will be influenced to a very minor extent—but nevertheless influenced—by all the planets and all the stars in all the galaxies. Secondly, what is "the correct" value of g_E? Since the Earth is not a perfect sphere, r, and therefore g_E, vary. There is another reason; the rotation of our Earth gives my falling chalk a centrifugal acceleration, so that g_E is dependent upon the latitude. And then . . . , but you can find others for yourself. Isn't it a wonderful thing that idealization enables effective investigation? For otherwise, surely Nature's complexity would bury her laws too deep for man to probe.

3.2.8 Hindsight and Foresight

Newton was not the only one, nor the first, to conjecture the Inverse Square Law. His brilliant scientific friends, Halley who on the basis of Newton's mechanics made with spectacular success the first prediction of a comet's return, Hooke who is remembered by his Law of Elasticity that the tension of a wire is proportional to its stretch, and Wren whose solid mathematical achievements are overshadowed by his architectural, all thought of it. The crucial difference is that they lacked that combination of insight and mathematical ability necessary to lock it in with Kepler's laws. Newton turned the key, his colleagues couldn't; they couldn't find a key to turn.

In retrospect, Newton's theory seems obvious. How could it possibly have been otherwise? Oh yes, told that this is the key to turn, and this the way to fit it into the lock, the rest is obvious. It is tempting to say that Halley, Hooke, and Wren also found the key—but wouldn't this be misleading, really? What use is a key if you can find no lock for it to fit?

Kepler also thought of the Inverse Square Law; he thought of it first. It is interesting to see how he arrived at it and especially interesting to see why he rejected it.

Kepler regarded gravitational attraction as analogous to propagation of light. His analogue is concerned with the intensity of propagation. Let us introduce this necessary preliminary.

It is an inescapable observation that the Sun emits light; without sunlight there would be no life on Earth. Climate is related to latitude, for

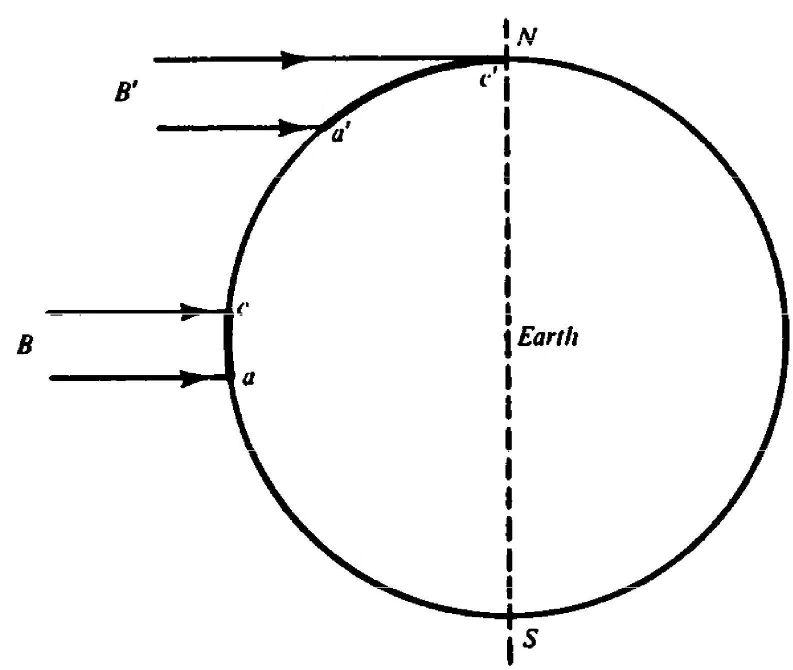

Figure 3.25

on latitude depends the angle at which the Sun's rays strike the Earth, and on the angle depends the area of the Earth's surface over which an incident beam of sunlight is distributed. See Fig. 3.25. Because of the Sun's great distance from the Earth, its beams, e.g., B, B', will be sensibly parallel. And supposing, as seems natural, that the Sun radiates light equally in all directions, beams B, B', if of equal breadth, will contain equal quantities of sunlight. Equal quantities may be distributed over unequal areas, for obviously area ac is less than $a'c'$. With the sunlight more thickly spread, more heating; the tropics are hotter than the poles. Thus the concept

$$\text{intensity of sunlight} = \frac{\text{quantity of light}}{\text{area}}$$

naturally presents itself. But, 20 units of light falling uniformly on 2 square centimeters is 10 units falling on each square centimeter, i.e.,

$$\text{intensity} = \text{quantity of light per unit area.}$$

It is not necessary for pursuit of Kepler's line of thought to consider in detail how the quantity of sunlight is to be measured or the unit to employ.

Consider now the intensity of light falling on a planet P at a distance R from the Sun. Let S be the total amount of light emitted by the Sun. Again, as seems natural, we suppose this to be radiated equally in all directions, so that the intensity will be the same at all points distance R from the Sun. But these points constitute a spherical sheet (with center the Sun) whose radius is R and whose surface area, therefore, is $4\pi R^2$. Consequently,

$$\text{intensity of radiation at } P = \frac{S}{4\pi} \cdot \frac{1}{R^2},$$

i.e., the intensity is inversely proportional to the square of the distance between the planet P and the Sun. See Fig. 3.26.

Since light is radiated from the Sun according to an inverse square law, could not gravitational attraction be similarly "radiated"? Kepler thought carefully about the possibility, but was dubious—and so missed a great discovery. That he did so, or rather that he was dubious, is to his credit; he mistrusted the idea for a very good reason. His reason? That although during a solar eclipse the Moon blocks the Sun's radiation to part of the Earth, there is no discontinuity in the Earth's motion. If gravitational attraction were radiated *as light is radiated*, this too would be temporarily blocked by the Moon, so that during the eclipse it would discontinue its

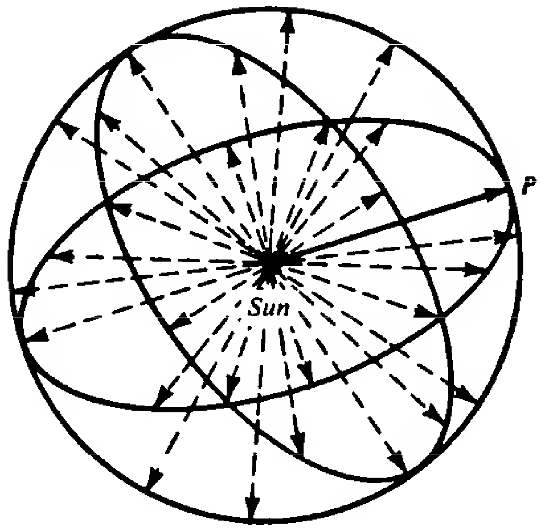

Figure 3.26

elliptical orbit about the Sun. But, it doesn't. Therefore, gravitational attraction is not radiated as light is radiated. See Fig. 3.27.

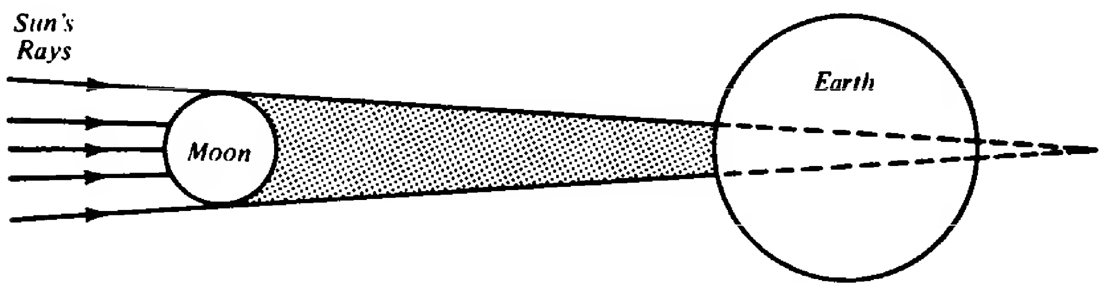

Figure 3.27

Kepler's argument is a good argument; do not let hindsight prejudice you against its merits: any fool can be wise after the event. Try to view the problem anew. That the Sun can keep the Earth in orbit and the Earth the Moon without any material connecting link is, to say the least, a most mysterious affair. It has been calculated that a steel cable, equal in cross section to the Moon's diameter, would not, fastened from Earth to Moon, be strong enough to replace the Earth's gravitational pull on the Moon. How can empty space be stronger than steel? The notion has the improbability of H. G. Wells' Time Machine.

Only a lunatic or a genius could believe the Moon to be kept in orbit by a force transmitted through empty space. Kepler, not a lunatic, rightly rejected the inverse square conjecture; Newton, not a lunatic, rightly accepted it. Kepler's conclusion was right relative to his partial understanding (and partial misconception) of the problem. Newton was justified for he—if you will pardon the colorful phrase—was not misled by that red

herring, white light. He saw clearly what Kepler could not appreciate; that consequent upon Galileo's Law of Inertia, an orbiting body must have an acceleration towards the concave side of its path—and we are back to the falling apple, the cannon ball, and the Moon.

The rest of the story we know—Newton's eager utilization of Kepler's three laws. Yet it would be a mistake to suppose that Kepler's works displayed his laws for the convenience of posterity; chameleon-like they were camouflaged by their context. Kepler, the last in the great Pythagorean tradition, had the magnificent ambition to explain the whole universe, lock, stock and barrel, in one devastating, all-embracing synthesis of geometry, music, astrology, astronomy and epistemology. Newton was less ambitious. In Kepler's *Harmony of the World* (1618), the sequel to his *Cosmic Mystery* (1597) and culmination of his lifelong obsession to establish the harmony of the spheres (for details, again see Koestler's *The Watershed*), his laws are a small part of the flotsam and jetsam cast up by his restless tides of thought. It remained for Newton to pick over the driftwood. He was a beachcomber of genius.

And how shall we best remember him? His friend Sir Christopher Wren, architect of St. Paul's Cathedral and hosts of other famous buildings, was fond of saying, "If you want to see my monuments, look around you." Were Sir Christopher alive to make that remark to Sir Isaac today, one can well imagine the latter's retort; a shrug of the shoulders followed by a sly jerk of the head in the direction of the Sputniks, Luniks, Pioneers, etc. Daily, his monuments become more numerous.

SECTION 3. THE PENDULUM

Primarily for two good reasons we begin this section with the common or garden variety of pendulum such as makes any grandfather's clock go tic-toc, tic-toc, and was used by Galileo in his experiments considered earlier: firstly, because derivation of the right kind of formula for its period of oscillation is the classic illustration of the dimensions test; secondly, because this formula is essential to verification of Newton's Law of Universal Gravitation by pendulum determination of g.

3.3.1 The Dimensions Test

This has nothing to do with the Hollywood producer's measure of a female filmstar's probable box office appeal; it is a test to ensure that formulae make sense, that the quantity indicated by the left-hand side of an equation is of the same kind or category as, or is syncategorious with, that indicated by the right.

For example, suppose it conjectured that the volume v of a sphere is given by

$$v = c \cdot r^2$$

where r is the radius of the sphere and c a specific (but here unspecified) number independent of r. Since r is a length, r^2 is an area, and $c \cdot r^2$ a larger or smaller area than r^2 according as c is greater or less than unity. In short, the formula states that a volume is identical with an area, or that a quantity measurable in cubic units is the same as a quantity measurable in square units. Isn't this an absurd thing to say? The quantities are not syncategorious, they are of different kinds.

Contrast this formula for v with

$$v = c \cdot r^3.$$

Here v and $c \cdot r^3$ are both measured in cubic units, so that the quantities are of the same kind and therefore comparable. The formula is the right kind of formula, it makes sense. If $c = 4\pi/3$, we have the right formula of the right kind; yet note that if c is taken to be any other number, say 16π, the formula still makes sense. It happens to be false; we have a wrong formula of the right kind. Conceivably c could have been 16π; what is inconceivable is that, for example,

$$v = 16\pi \cdot r^2$$

or

$$v = \tfrac{4}{3}\pi \cdot r^2.$$

It just doesn't make sense to say that a volume is equal to an area; the quantities are not syncategorious, they cannot be compared. Likewise, the teen-ager who says that his age is fourteen hundredweights has confused his categories.

The dimensions test is the basic logical grammar of physics. Although, unfortunately, it does not ensure that our equations must be true, it does ensure that they do make sense, that they could conceivably be true. The dynamicist's working concepts, for example, velocity, acceleration, force, impulse, work, momentum, energy, power, may each be defined in terms of (at most all three of) the basic concepts or dimensions, length, mass, and time. For example,

$$\text{kinetic energy} = \tfrac{1}{2}mv^2,$$

where, of course, m is the mass and v the velocity in question. But, velocity is defined as displacement or length per unit time. Taking, as is usual, the letters L, M, and T for length, mass, and time, velocity is indicated schematically by

$$\frac{L}{T} \text{ or } L \cdot T^{-1}.$$

It has 1 dimension of length, -1 dimension of time—and, if you wish to be fussy, 0 dimension of mass, for its dimensions could be indicated schematically by

$$L^1 \cdot M^0 \cdot T^{-1}.$$

Consequently, proceeding schematically, for v^2 we have

$$(L^1 \cdot M^0 \cdot T^{-1})^2 \text{ or } L^2 \cdot M^0 \cdot T^{-2}$$

and, for kinetic energy

$$(\tfrac{1}{2}M) \cdot L^2 \cdot M^0 \cdot T^{-2}$$

or, respecting the alphabet and ignoring the pure number $\frac{1}{2}$ (since this affects only the amount of the quantity considered, not its quality or category),

$$L^2M^1T^{-2}.$$

Next, let us check the dimensions of (16), the formula so important for Newton's verification. Since acceleration can be measured in cm/sec^2 the dimensions of g_E may evidently be indicated schematically by

$$\frac{L}{T^2} \text{ or } L^1 \cdot T^{-2}.$$

But, proceeding schematically,

$$\frac{4\pi^2R^3}{T^2r^2} = (\text{Number}) \cdot \frac{L^3}{T^2 \cdot L^2} = \frac{L}{T^2} \text{ or } L^1T^{-2}.$$

Check. The test does not show that the formula is the correct formula, but it does show that it is the right sort of formula, that it makes sense. Had the test failed the formula could not have made sense, it would have

been absurd. The test is a necessary but not sufficient condition for correct formulae.

Let us sum up: Any physical quantity Q has basic dimensions α, β, and γ of length, mass, and time (and there are no others). Schematically

$$Q = L^{\alpha}M^{\beta}T^{\gamma}.$$

And, if

$$Q' = L^{\alpha'}M^{\beta'}T^{\gamma'},$$

Q and Q' are quantities of the same kind (but not necessarily of the same amount) if and only if

$$\alpha = \alpha', \quad \beta = \beta', \quad \text{and} \quad \gamma = \gamma'.$$

3.3.2 Simple Pendulum's Time of Swing

As with Galileo's pendulum experiments, we suppose idealization, that the frictional resistance of the air, the weight of the string, and the dimensions of the bob may be neglected. It is of course essential that the bob be heavy; with a feather for a bob air resistance is obviously not negligible. We suppose the length of the string to be l. See Fig. 3.28.

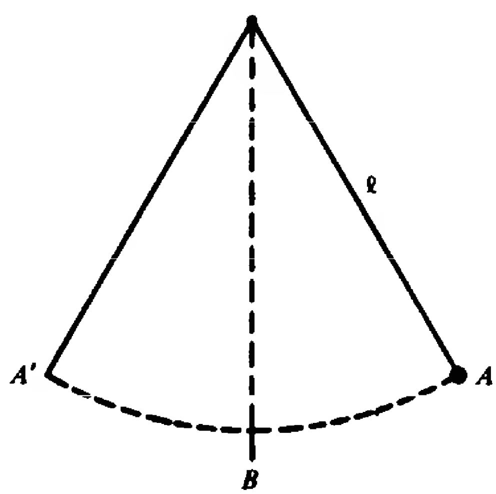

Figure 3.28

Our problem is to find T, the time of swing, or more solemnly put, period of oscillation, of the bob. By this we mean the time of a complete swing, from A across to A' and back to A, as when a grandfather's clock goes tic-toc-tic. This point is of importance as many textbooks give a formula for only a half-swing; tic-toc without the succeeding tic.

On what does T depend? We must make a conjecture; to jug hare, first catch your hare.

Let us begin with what we know, congenital physics. Isn't there some analogy between the swing of a pendulum and the swing of the leg as one strides along? Even in this car-ridden country motorists become pedestrians to reach their automobiles. Haven't you ever lingered at a street corner to observe how people walk? Presumably Aristotle did—for he observed that there is a minimum speed at which one can walk. Also there is a comfortable speed for each pedestrian, when the swing is natural and unforced; short legs naturally swing more quickly than long ones. Doesn't this suggest that a pendulum's time of swing depends upon the length of its leg?

The simplest assumption, that T is directly proportional to l, is disproved by a minimum of experiment. Yet T clearly depends on l, so what is the next simplest conjecture? Let us suppose that T is proportional to some power of l, say l^{α}.

On what else does T depend? Does it depend on the mass of the bob? By experiment (keeping l constant) we find that provided the bob is heavy, thereby keeping air resistance relatively small, it does not matter how heavy.

What else? If there were no gravitational field the pendulum would not swing at all. So, supposedly under a very weak gravitational field it would swing to and fro ever so slowly. Doesn't it seem reasonable to suppose that as g becomes greater, T becomes smaller? But the dependence need not be simple inverse proportion; so let us suppose that T is proportional to g^{β}, where β is expected to be negative.

Thus we have grounds for conjecturing that T is proportional to l^{α} and to g^{β}, but independent of the mass of the bob; i.e., that

$$T = cl^{\alpha}g^{\beta}.$$

Have we taken all the relevant factors into account? Not being able to think of any others, let us apply the dimensions test to this equation.

Schematically, for the left-hand side we of course have

$$T = L^0M^0T^1.$$

And for the right-hand side? c is a pure number and only affects the amount, not the quality, and so may be ignored. g can be measured in cm/sec^2, so that its dimensions, as we ought to expect from Galileo's

work, are those of acceleration, LT^{-2}. So, schematically,

$$\begin{aligned} cl^{\alpha}g^{\beta} &= L^{\alpha}(LT^{-2})^{\beta} \\ &= L^{\alpha}(L^{\beta}T^{-2\beta}) \\ &= L^{\alpha+\beta} \cdot T^{-2\beta} \end{aligned}$$

which, making fully explicit that the formula is independent of the mass of the bob, we write as

$$L^{\alpha+\beta}M^{0}T^{-2\beta}.$$

Consequently, T and $cl^{\alpha}g^{\beta}$ have (as stated in the last paragraph of Number 3.3.1) the same dimensions of length only if

$$0 = \alpha + \beta$$

and the same dimension of time only if

$$1 = -2\beta$$

(they already have the same dimension of mass, 0). From the latter equation

$$\beta = -\tfrac{1}{2},$$

negative as we anticipated, and from the former,

$$\alpha = +\tfrac{1}{2}$$

giving

$$T = cl^{1/2}g^{-1/2}$$

i.e.,

$$T = c\sqrt{\frac{l}{g}}\,. \tag{17}$$

Oh yes, our conjecture was daring—yet there was nothing worse at stake than the possibility of being wrong and having to think again. As it happens our conjecture was fortunate. It remains to determine the numerical value of c.

Many dynamicists of ability tackled with unsuccess the problem of a formula for T. Galileo came close to solving it, yet never quite succeeded. Its complete solution demands use of differential equations. Finally it was deduced with less than full rigor by Huygens; his working knowledge of the calculus was not quite adequate for a fully explicit derivation. It turns out that $c = 2\pi$, and that the formula is accurate only if the oscillations are small. It would not do, for example, to have the pendulum swing through half circles, but when the pendulum string does not oscillate more than a few degrees from the vertical the formula is quite accurate, even for scientific purposes. So, for small oscillations,

$$T = 2\pi\sqrt{\frac{l}{g}}\ . \tag{18}$$

3.3.3 Determination of g by Pendulum Experiment

An explicit formula for g is immediately available. We square (18),

$$T^2 = 4\pi^2 \frac{l}{g},$$

so that

$$g = \frac{4\pi^2 l}{T^2}\ . \tag{19}$$

Thus g (for here the subscript of g_E in (16) may be dropped without confusion), so essential to the verification of Newton's theory, may be obtained experimentally by observation of a pendulum's period of oscillation.

The accurate measurement of l is no problem, but how is T to be measured accurately? Obviously, to take, say, one hundredth of the time of one hundred complete oscillations is better than to time a single oscillation, for then the error in timing is, so to speak, dispersed over a hundred individual observations. And to determine the bob's return to a former position it is best viewed against the hairline of a telescope. It would not do to set the hairline at A (see Fig. 3.28 again), for air resistance, though small, does have a damping effect on the oscillations, so that they gradually become smaller. The obvious position for the hairline is B, along the vertical through the bob's point of support; for the motion, as we know from Galileo's experiments, is symmetrical about it.

3.3.4 The Conical Pendulum

The determination of T, the period of oscillation of a conical pendulum, is a somewhat similar problem to the determination of T for the simple pendulum, but has the advantage that we can solve it completely even with the mechanics of the preceding lectures.

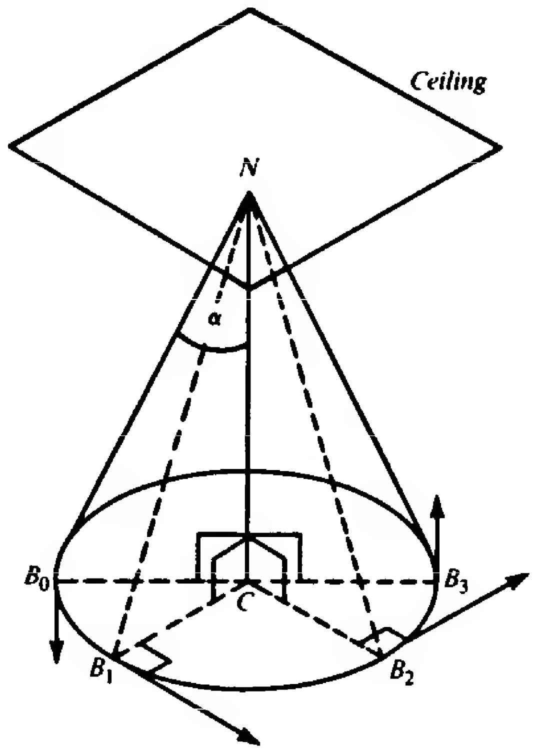

Figure 3.29

First, what is a conical pendulum? The apparatus is that of the simple pendulum; the distinction lies in the path of the bob. If the bob swings to and fro on an arc of a vertical circle, the pendulum is simple; if it rotates in a horizontal circle, the pendulum is conical. Suppose a bob B to be suspended from a nail N in the ceiling by a string NB, and let C be the foot of the (vertical) normal in the horizontal plane through the initial position B_0 of the bob. See Fig. 3.29. Let B start from rest, and it will swing to and fro in an arc from B_0 of a circle of center N and radius NB_0 in the vertical plane NB_0C: we have a simple pendulum. Alternatively, give B an appropriate push in the horizontal direction perpendicular to B_0C and (supposing the string free to swivel at N) it will rotate in a circle of center C and radius B_0C in the horizontal plane through B_0: we have a conical pendulum. As B circles C, NB generates the lateral surface of a cone; the term conical is appropriate.

As already mentioned when speaking of Galileo, as well as being simple the apparatus is inexpensive. Here is another experiment you can perform for an infinitesimal outlay—provided that you already have a roof over your head. Moreover, in contrast to make-your-own-atomic-pile experiments, there is no risk of burning the house down or blowing the ceiling up; although not destructive, instructive.

Yet even without performing any fire-proof experiments we already know, willy-nilly, something of the results by congenital mechanics. Suppose the horizontal circle of rotation of B to be such that NB is inclined at an angle α to the vertical NC, i.e., such that α is the semi-vertical angle of the cone generated by the string. If α is increased (with the length of the string invariant) will T, the time of revolution of the bob, increase or decrease? Part of the answer we can feel with our muscles; the nearer B's plane of revolution to the ceiling, the greater the thrust to get it started. Could we get B to rotate in the plane of the ceiling? Can't you feel the situation? No, not quite in the plane of the ceiling. To get B to rotate in a plane just beneath the ceiling would require an enormous initial thrust. Don't your muscles ache at the very thought of it? And with an enormous initial thrust B would rotate with enormous velocity. So? We conclude that the nearer α becomes to 90°, the nearer T becomes to 0. T depends on α.

On what else does T depend? Consider a limiting case. If, in the complete absence of a gravitational field, B is started on a horizontal circular path at the ceiling, it will continue on this path for there is no force to pull it down. So shouldn't we expect T to depend on g?

Next suppose that α is kept constant and that l, the length of the string NB, is increased. When l is increased, the radius of B's horizontal circle of rotation is increased. With a short string, just an inch or two long, would B need a smaller or a greater push than if the string were several feet long? Your muscles tell you that the longer the string, the greater the force. The greater the force, the greater the velocity. But there is a complication; the longer the string, the greater the circumference of the circle of oscillation. So, the farther the bob has to go to complete an oscillation, the faster it goes. Does the increase in velocity more than compensate or less than compensate for the increase in distance? Does T decrease as l increases, or does it increase? What seems unlikely is that the increase in velocity with increase in l *just* compensates for the increase in distance, thereby making T independent of l.

In short, it would appear *prima facie* that T depends on l as well as on g and α. Recalling the formula for T for the simple pendulum and

remembering the dimensions test, is it really too wild a conjecture that for the conical pendulum

$$T = c \cdot \sqrt{\frac{l}{g}} \cdot f(\alpha),$$

where $f(\alpha)$ is of zero dimensions? Be this as it may, it would appear that a correct formulation of our problem is:

Given α, g, l, *find* T.

Take another look at Fig. 3.29. Isn't it a matter of complete indifference whether we start the bob rotating at B_0, B_1, or B_2? Whatever forces are acting on B initially will continue to act and no new forces are introduced. By the Law of Sufficient (or Insufficient) Reason there is no reason why the motion should not be uniform. Therefore B has uniform circular motion. We recall Newton's argument for central acceleration so that B must have a centripetal acceleration a, and that, according to Hamilton's hodograph deduction,

$$a = \frac{v^2}{r},$$

where v is the initial horizontal velocity perpendicular to B_0C, and $B_0C = r$.

But how is this acceleration caused? g acting vertically (downwards) has of itself no horizontal component; there must be a second force. If N is not securely hammered into the ceiling, it will be wrenched out by the motion. There is a tension in the string. Thus a is the resultant of two forces acting on B; g vertically downwards and the tension in the string obliquely upwards. We complete the parallelogram of forces: See Fig. 3.30. From the obvious geometry of the figure $BG = N'C'$, so that considering the right $\triangle BN'C'$, we have

$$\tan \alpha = \frac{a}{g}.$$

We have related a to the geometry of the figure.

Our problem, remember, is to find T—supposedly in terms of α, l, and g. So far, we do not have what we must have—an equation containing T, and we do have what ultimately we must not have—the intrusion of v and r. We must introduce T and eliminate v and r. Can we kill two birds with one stone? Since the motion is uniform circular

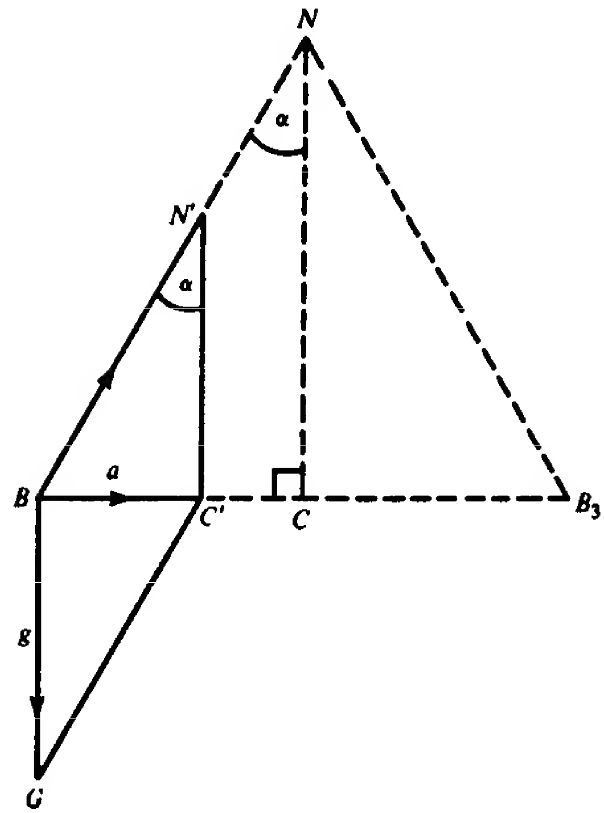

Figure 3.30

motion, $2\pi r$ being the circumference of the circle of oscillation of radius r, we have by definition that

$$v = \frac{2\pi r}{T}.$$

Do we now have enough equations to find T? The business of solving problems is so important to mathematicians and the business of getting into formulae problems stated in words so important to many who are not primarily mathematicians (such as engineers and chemists) that it is well worthwhile to emphasize by tabulation the role of our equations and how we obtained them. We tabulate:

$a = \dfrac{v^2}{r}$	from dynamics: the principal equation
$\tan \alpha = \dfrac{a}{g}$	by geometry of vector triangle
$v = \dfrac{2\pi r}{T}$	since v is constant

How many quantities are there? Six, namely, a, v, r, α, g, and T. How many of these are known? Two, namely, α and g. So that leaves four unknowns, namely, a, v, r, and T. Primarily we are interested in T. The others, a, v, and r, are only means to an end, so let us term them auxiliary unknowns; their role is to help solve the problem. Yet characterizing their role will not alter the fact that whereas we do have four unknowns we have only three equations. We need a fourth equation.

What quantities ought the fourth equation contain? In mathematics, if not in metaphysics, it is necessary to be clear about what you are doing. Again we recall the original problem:

Given α, g, l, *find* T.

Note that our above list of six quantities does not contain l. Although l is given, it has not been taken. So look at NB in Fig. 3.30. An obvious relation involving l is

$$\sin \alpha = \frac{r}{l}.$$

We have obtained a fourth equation which introduces l without introducing any new unknowns. Four equations, four unknowns; the stage is set for the determination of T.

It is my personal opinion that there is nothing of greater importance to be taught in mathematics to the high school student than the business of setting up equations. For better or for worse, whether we like it or not, we live in a technological society that daily becomes increasingly so. Although a typical student will not become a professional mathematician, the nearer his future field of endeavor to science, the greater his need to understand textbooks, manuals, journals, and articles in which mathematical formulae are steadily becoming more numerous. And unless he is to rest content throughout the whole of his life as a non-contributor to his chosen field, he will at least need be able to set up similar equations for himself. It is not in the nature of intelligent man to be a spectator to life.

To repeat a point whose importance in my view justifies its repetition: without getting to understand what a problem is about and what is relevant to it and (when appropriate) translating it from words into formulae, there is no mathematical education. That most of the word problems in the traditional textbooks are so boring and useless does not invalidate my point: of course, the problems must be intelligently designed. Is it not significant that even at a time when technology was attending an antenatal clinic, Newton, Euler, and Descartes each thought the topic of solving "word problems" and the setting up of equations worth his writing about?

We return to the determination of T. From the second tabulated equation,

$$g \tan \alpha = a.$$

Substituting for a in the first equation, we get

$$g \tan \alpha = \frac{v^2}{r}.$$

Squaring the third equation and substituting for v^2, we have

$$g \tan \alpha = \frac{4\pi^2 r^2}{T^2} \cdot \frac{1}{r} = \frac{4\pi^2}{T^2} \cdot r,$$

so that

$$T^2 = \frac{4\pi^2}{g \tan \alpha} \cdot r.$$

Finally, from the fourth equation,

$$r = l \sin \alpha,$$

so that

$$T^2 = \frac{4\pi^2}{g \cdot \tan \alpha} \cdot l \sin \alpha = 4\pi^2 \cdot \frac{l \cos \alpha}{g}$$

giving

$$T = 2\pi\sqrt{\frac{l}{g} \cdot \cos \alpha}\ . \tag{20}$$

That (20) is of the form

$$T = c\sqrt{\frac{l}{g}} \cdot f(\alpha),$$

where $f(\alpha) = \sqrt{\cos \alpha}$ and is of zero dimensions confirms our conjecture.

A consequence of (20): as α tends to $90°$, $\cos \alpha$ tends to 0, and consequently so does T. For the bob to rotate in the plane of the ceiling its velocity would need be infinite. Mathematical deduction supports muscular perception.

And the other limiting case? As α tends to 0, $\cos\alpha$ tends to 1, and consequently T tends to $2\pi\sqrt{l/g}$. Thus, most curiously, when the circle of oscillation becomes very small, the period is the same as that of the simple pendulum.

In conclusion, note the importance of the conical pendulum for Newton's Theory. Here is a simple demonstration of the necessity for a centripetal acceleration for uniform circular motion, a limiting case of planetary motion. Reconsider the circumstances of Fig. 3.30 as in Fig. 3.31. The force of gravity g acting on B may be decomposed into a force $\overrightarrow{BN}$ acting along NB and a force $\overrightarrow{BC'}$ along BC, as is indicated by the parallelogram of forces. The first component is utilized in keeping the string taut. What about the "spare" force along BC? This provides the centripetal acceleration necessary for uniform circular motion.

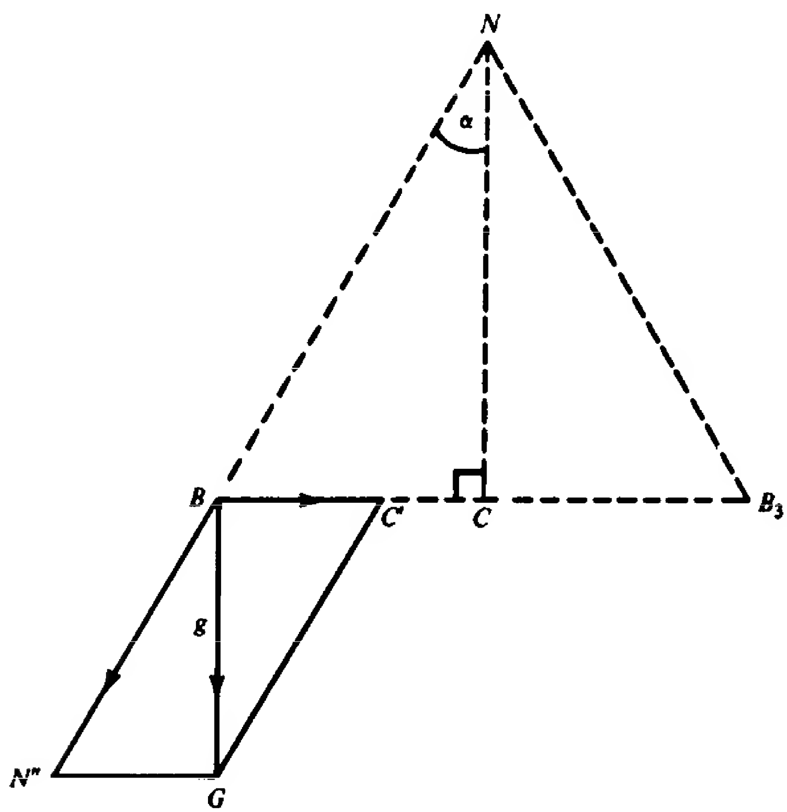

Figure 3.31

SECTION 4. ESCAPE VELOCITY

This section, despite its title, is not about the rate of departure of a convict over the penitentiary wall. Our concern is much more exciting; namely, the velocity necessary for a space capsule to excape the Earth's gravitational pull. Although convicts have been escaping from penitentiaries for as long as there have been penitentiaries to excape from, it is

only within the last few years that technological progress has made possible generation of the extremely high velocities to put satellites into orbit about the Earth and to send rockets to the Moon and to outer space. What was science fiction is rapidly becoming fact; the other side of the Moon has been photographed. The space race is supplanting the World Series in public interest; we live in the Satellite Age.

With ever increasing frequency projectiles are being hurled into space. Students, stimulated by newspaper, radio, and television reports, will be eager to know more. Better students will ask better questions; among others, questions bearing on the relevance of mathematics to space travel. The answer to many such questions is a difficult complex of differential equations, but happily there are some such topics of Newtonian mechanics amenable to elementary treatment. An especially amenable topic is the velocity of space capsules, so let us consider it.

We begin by distinguishing between orbital and escape velocity; by the former we mean the velocity of a satellite in orbit about the Earth, by the latter the velocity necessary to escape from the Earth's gravitational field to outer space. Alternatively, we may use the self-explanatory terms *go-around* and *go-away* velocities. See Fig. 3.32. Which velocity do you suppose the greater? Many persons reply "The go-away velocity, obviously." Here "obviously" is all too apt to mean groundless conviction. You have grounds for your conjecture? Either way you have committed yourself; the question must be answered. Since the go-around velocity is easier to compute, let us deal with it first.

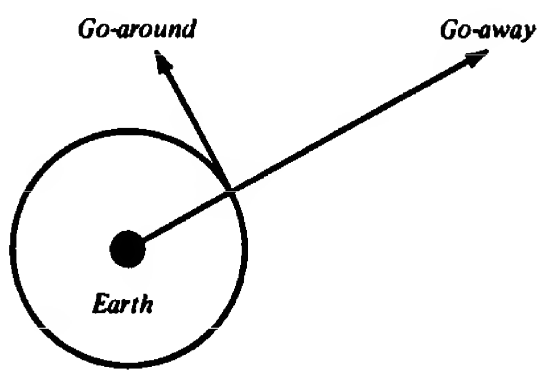

Figure 3.32

3.4.1 Go-Around Velocity

Suppose a satellite in orbit about the Earth, just skimming the chimney pots. Not a very realistic supposition; very dangerous for the tree tops. Moreover, although we do not feel the friction of the air when strolling along, the friction of the Earth's atmosphere is tremendous at high velocities and horribly complicates our problem. So, idealization is imperative; we suppose no atmosphere. Also, that the orbit will be precisely a

circle instead of an ellipse and that the velocity will be exactly uniform. Now the main idea, of course, is that the Earth's gravitational attraction must provide our satellite with a centripetal acceleration. We recall Hamilton's hodograph derivation of centripetal acceleration, namely (11),

$$a = \frac{v^2}{r}.$$

What, for our present problem, is r? Since our satellite is skimming the roof tops, r is the radius of the Earth. And what is a? a is the acceleration due to the Earth's gravitational pull. But the strength of this pull depends upon the distance of our satellite from the Earth. To be precise a is the acceleration due to the Earth at the Earth's surface. To remind ourselves of this, as in an earlier context, we denote this acceleration by g_E rather than g. (E is for *Earth* and for *Emphasis*.) We have

$$g_E = \frac{v^2}{r},$$

so that

$$v = \sqrt{g_E \cdot r}\,. \tag{21}$$

We have found the go-around velocity at the Earth's surface.

More realistically, let us consider a satellite in circular orbit, say, 300 kilometers above the Earth's surface. At this height our satellite is above the Earth's atmosphere, so that friction, if any, is negligible. We can immediately write that v_{300}, the go-around velocity at 300 kilometers above the Earth's surface, is given by

$$v_{300} = \sqrt{g_{300} \cdot r_{300}}\,,$$

where g_{300} is the acceleration due to the Earth's gravitational attraction 300 kilometers above its surface and r_{300} the orbit radius (300 kilometers more than the radius of the Earth). The problem to find v_{300} is reduced to the problem to find g_{300}.

The latter problem can be avoided by using Kepler's Third Law, yet another fact that underlines the crucial role of this law in Newtonian dynamics. Let T be the orbital period at the Earth's surface and T_{300} the period 300 kilometers above it. Since the motions are supposedly uniform circular motions,

$$T = \frac{2\pi r}{v}, \qquad T_{300} = \frac{2\pi r_{300}}{v_{300}},$$

where $r_{300} = r + 300$, r being measured in kilometers. Hence

$$\frac{T}{T_{300}} = \frac{2\pi r}{v} \cdot \frac{v_{300}}{2\pi r_{300}};$$

i.e.,

$$\frac{T}{T_{300}} = \frac{v_{300}}{v} \cdot \frac{r}{r_{300}}.$$

At this stage we make use of Kepler's Third Law

$$\frac{T}{T_{300}} = \frac{r^{3/2}}{r_{300}^{3/2}}.$$

From this last pair of equations

$$\frac{v_{300}}{v} \cdot \frac{r}{r_{300}} = \frac{r^{3/2}}{r_{300}^{3/2}},$$

so that

(22) $$v_{300} = \sqrt{\frac{r}{r_{300}}} \cdot v,$$

and from the first of the pair

(23) $$T_{300} = \left(\frac{r_{300}}{r}\right)^{3/2} \cdot T.$$

We leave as an exercise calculation of the go-around velocity and period of space capsules in orbit 100, 200, 300 and k kilometers above the Earth's surface. Surely the reader will be keen to work out in this way the average go-around velocity and approximate period of any orbits actually being made, to compare their answers with the figures publicly announced in newspapers and on radio and television.

3.4.2 Apropos Go-Away Velocity

As already remarked, the calculation of the go-away velocity is more difficult. More difficult, because the Earth's gravitational pull on a rocket heading for outer space is not constant, but varies with the rocket's distance from Earth. The answer depends upon the instantaneous deceleration at every point of its path from the Earth's surface to outer space. In

calculating the go-around velocity v_{300} we were able to avoid determination of g_{300}; in determining the escape velocity we cannot avoid knowing the various g's. Yes, the problem is more difficult.

What is the Earth's gravitational pull at a distance x kilometers from its center? According to Newton's Law of Universal Gravitation the pull gives our rocket a deceleration inversely proportional to the square of the distance; i.e., proportional to $1/x^2$. Hitherto, we have been concerned with the effect of gravity, acceleration or deceleration; it is only indirectly that we have been concerned with its cause, the force of gravity. In our next problem it is convenient to deal with gravity in terms of the force it exerts rather than in terms of the change of velocity it effects. Accordingly we now consider this necessary preliminary.

3.4.3 The Force of Gravity

Let us suppose that you are weekend climbing the mountains of the Moon. There, as here, your rucksack contains spare socks and a pint flask of brandy for medicinal purposes. Although on the Moon your spare socks are still the same size and the brandy (before the emergency) still fills the flask, each article weighs less; the Moon's gravitational pull is about one-sixth that of the Earth's. Whereas mass remains the same, the force to which it is subjected does not. The weight of the mass or substance is the measure of the force exerted on it. If on Earth the force exerted by gravity on a pint flask of brandy is 1 pound weight, on Moon it weighs about $\frac{1}{6}$ pound; in either place two pint flasks of brandy weigh twice as much as one does. If when climbing in either place your flask is let slip, you can console yourself with the thought that two would have fallen just as fast. Although two weigh twice as much as one, although the gravitational force exerted on two is twice that exerted on one, the accelerations are the same. A force of $2m \cdot g_E$ acting on a mass $2m$ produces in it an acceleration g_E as does a force $m \cdot g_E$ acting on m; a force $2m \cdot g_M$ acting on a mass $2m$ produces in it an acceleration g_M as does a force $m \cdot g_M$ acting on m. This conception of force, mass and acceleration is embodied in Newton's Law

$$\text{force} = \text{mass} \times \text{acceleration}. \tag{24}$$

And we recall that Kepler's laws are valid for Jupiter's moons as well as for the Sun's satellites, so that as a consequence of his third law the centripetal acceleration of each of Jupiter's moons towards Jupiter is inversely proportional to the square of its distance from Jupiter, just as (in consequence of Kepler's Third Law) the centripetal acceleration of each of the Sun's satellites towards the Sun is inversely proportional to the

square of its distance from the Sun. Likewise, the Moon's centripetal acceleration is inversely proportional to the square of its distance from the Earth. The moons or satellites of each system each have a centripetal acceleration such that

$$\text{centripetal acceleration proportional to } \frac{1}{(\text{distance})^2}$$

or

$$\text{centripetal acceleration} = \frac{c}{(\text{distance})^2}\,. \tag{25}$$

The point to note is that whereas c, the constant of proportionality, is the same for all the moons or satellites of the same system, it is different for different systems. How did Newton get a truly universal law, a law in which the constant of proportionality is the same for all systems?

Combination of (24) with (25) enables us to consider gravity in terms of the force it exerts instead of in terms of the acceleration it causes. Combining these equations, we see that a planet's gravitational pull on its satellite is directly proportional to the satellite's mass and inversely proportional to the square of the satellite's distance. Let F_m be the force exerted on a satellite of mass m to cause it an acceleration g at a distance r from its attracting body. By (24)

$$F_m = m \cdot g,$$

and by (25)

$$g = \frac{c}{r^2},$$

so that

$$F_m = c \cdot \frac{m}{r^2}\,.$$

Note well that the attracting force is a function of the mass of the attracted body as well as its distance.

The Moon is attracted by the Earth, the Earth is attracted by the Sun. It is natural, but audacious, to conjecture that every particle is attracted by every other, that gravitation is universal. But, if particle A attracts particle B and every particle attracts every other, then also B attracts A. If every particle attracts every other, then every particle is attracted by every other.

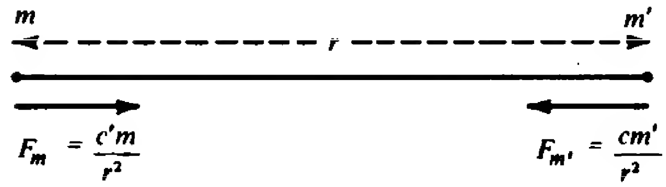

Figure 3.33

Consider the simplest universe of this kind, namely that with just two particles, say m, m', distance r apart. See Fig. 3.33. If m is regarded as a "satellite" in the gravitational field of m', then the force exerted on it is directly proportional to its mass m and inversely proportional to the square of its distance from m'; i.e.

$$F_m = c' \frac{m}{r^2}, \tag{26}$$

where c' is a constant of proportionality for m''s gravitational field. According to Newton's theory, this constant c' of proportionality depends only on m'. What is the nature of this dependence? To answer this question, we now regard m' as a satellite in the gravitational field of m. The force $F_{m'}$ exerted by m on m' is then given by

$$F_{m'} = c \frac{m'}{r^2}, \tag{26'}$$

where the constant of proportionality c, associated with m's gravitational field, depends only on m. Because of the complete symmetry of the situation, the relation of c to m is the same as that of c' to m'.

Now according to another law of Newton's, the law of action and reaction, the forces two bodies exert on each other are the same in magnitude and opposite in direction:

$$F_m = F_{m'}.$$

Then by (26) and (26')

$$\frac{c'm}{r^2} = \frac{cm'}{r^2},$$

so that

$$\frac{c}{m} = \frac{c'}{m'}.$$

Since c depends only on m, c' only on m', the left side is independent

of m', the right independent of m. We denote the common value of these ratios by G:

$$\frac{c}{m} = \frac{c'}{m'} = G;$$

and since m and m' are arbitrary pairs of masses, we conclude that G is independent of either mass. When we substitute $c = mG$ and $c' = m'G$ into (26) and (26′) we find

$$F_m = \frac{m'mG}{r^2}, \tag{27}$$

and

$$F_{m'} = \frac{mm'G}{r^2}. \tag{27'}$$

Each of these forces is directly proportional to m, directly proportional to m', and inversely proportional to r^2, with the *same* constant of proportionality G for the gravitational fields of *any* two particles m and m'. Being universal, G is appropriately termed the ***universal gravitational constant.***

Supposedly it is along some such line of thought that Newton arrived at (27), his Law for the Force of Gravitational Attraction, from his already established hypothesis that acceleration due to gravity is inversely proportional to the square of the distance.

To be precise, this law is taken by Newton to hold only for mass points, that is, for bodies whose dimensions are negligible. It is a consequence of this law—which Newton had considerable difficulty in proving—that the resultant attraction of a uniform sphere on a particle outside it is as if the whole of the mass of the sphere were concentrated at its center. So when dealing, for example, with the gravitational attraction of the Earth at a distance r from it, we must take r as the distance from its center, not its surface.

3.4.4 That Kepler's Third Law is a Consequence of Newton's Law of Gravitation

We recall that the crucial step in Newton's formulation of his Law of Universal Gravitation is that an inverse square law of gravitational attraction is a consequence of Kepler's law that the square of a planet's period T is proportional to the cube of its orbital radius r. We shall now show, conversely, that Kepler's Third Law is a consequence of Newton's.

With close approximation to the facts, we suppose a planet to move around the Sun with uniform circular motion. Let M and m be mass of Sun and planet, r the distance between them, and F the gravitational pull of the former on the latter. Any letters subsequently introduced are to be given customary interpretation. By Newton's Law of Gravitation, (27),

$$F = \frac{GMm}{r^2}.$$

And since the motion is uniform circular motion, the centripetal acceleration is given by (11),

$$a = \frac{v^2}{r};$$

and by (24), for the centripetal force F, we have

$$F = ma. \tag{28}$$

It remains to introduce T. Since the motion is uniform circular motion, by definition of velocity

$$v = \frac{2\pi r}{T}. \tag{29}$$

Our problem is to find the relation between T and r.

First we eliminate a. From (11) and (28),

$$F = m\frac{v^2}{r}.$$

Substituting in (27) to eliminate F, we have

$$m\frac{v^2}{r} = \frac{GMm}{r^2},$$

so that

$$v^2 = \frac{GM}{r},$$

and m also is eliminated. Squaring (29) and substituting for v^2, we have

$$\frac{4\pi^2r^2}{T^2} = \frac{GM}{r},$$

so that

$$(30) \qquad T^2 = \left(\frac{4\pi^2}{GM}\right)r^3.$$

$4\pi^2/GM$ is independent of T and r, so that

$$T^2 \text{ is proportional to } r^3,$$

as was to be shown.

The attentive reader will note that this deduction, as that in Number 3.3.4 to determine the period of oscillation of a conical pendulum, readily lends itself to detailed "Word Problem" development: What is given? What is to be found? How many equations?

3.4.5 Planetary Mass

From (30), solving for GM, we have

$$(31) \qquad GM = 4\pi^2 \cdot \frac{r^3}{T^2}.$$

The product of G and the mass M of the Sun is a function of r and T, the orbital radius and period of a satellite, but is independent of this satellite's mass. And of course the formula is equally applicable to any other sun and a satellite of that sun. Jupiter has moons, so let us apply it to Jupiter and one of its moons. If m is the mass of Jupiter and r' and T' the orbital radius and period of a moon of Jupiter, we have

$$Gm = 4\pi^2 \cdot \frac{r'^3}{T'^2},$$

so that

$$\frac{GM}{Gm} = \frac{4\pi^2 r^3/T^3}{4\pi^2 r'^3/T'^3},$$

i.e.

$$(32) \qquad \frac{M}{m} = \frac{(r/r')^3}{(T/T')^2}.$$

Thus we can compute the ratio of the masses of the Sun and Jupiter. Similarly, since the Earth has a moon, the ratio of the Sun's mass to the

Earth's can be determined, and consequently, the relative masses of Sun, Jupiter, and Earth.

To find M, the actual mass of the Sun, (31) is of course insufficient; we need to know G. From (27), when $M = 1$, $m = 1$, $r = 1$, we have

$$F = G.$$

That is to say that G is the gravitational force exerted by unit mass on unit mass at unit distance. So, in principle, G could be determined by measuring the force of attraction between two bits of blackboard chalk. However, the force is so very small that it is rather impractical to determine it in this way. G was determined with accuracy by the English physicist Cavendish, the German physicist von Jolly, and others. I wish to mention that one method of determination, using the torsion balance, was devised by the Hungarian physicist Eötvös—a brilliant lecturer and my teacher at the University of Budapest.

Actually (30) is not exactly correct. Since it makes use of the formula for the centripetal acceleration of uniform circular motion (Hamilton's hodograph result), it presupposes that the center of attraction is fixed. But the Sun is not nailed to a point in space; it moves. When you jump into the air you kick away from the Earth and so move it. True, such a small impulsive thrust acting on such a gigantic mass moves it a negligible distance. Although your prancing does not cause an earthquake, the distinction is correct. Taking the Sun's motion into account, which involves the concept of relative velocity, it turns out that (30) must be replaced by

$$T^2 = \frac{4\pi^2}{G(M+m)} \cdot r^3.$$

That is to say that the relation between T and r depends not only on the mass M of the Sun, but also on the mass m of the attracted satellite. And just to indicate that we have not exhausted the topic of planetary mass, let me conclude with the remark that it is possible to determine the mass of Mercury and Venus, even though they have no satellite moons. It is sufficient to know their orbital periods to the required degree of precision.

3.4.6 Go-Away Velocity

We arrive at the titular topic of Section 4: to determine the initial velocity with which a rocket must be fired from the Earth's surface to voyage straight on into outer space, never to return. With inevitable idealization we neglect the frictional resistance of the Earth's atmosphere.

Also, we suppose Earth and rocket to be the only bodies in the universe, so that the only force acting on our rocket is the Earth's gravitational pull. Even with these simplifications the answer is far from obvious.

What happens? Immediately our rocket blasts off, its initial velocity begins to decrease due to the Earth's gravitational pull. The farther away from the Earth the less the Earth's pull and the slower the rate at which our rocket's velocity decreases. Yet its velocity continually decreases and if it were finally to come to rest way out in space, its rest would be only momentary; the Earth's pull though weak, being the only force, would begin to draw our rocket back to Earth. The nearer our rocket returned to Earth, the greater the Earth's pull on it; it would arrive back at the speed with which it set out. The crux of the matter is to determine an initial velocity just sufficient to overcome the effect of the Earth's continuously decreasing gravitational pull.

How are we to set about computing this continuously variable effect? Let us cast our minds back to the section on Galileo. He dealt with the continually changing in terms of the unchanging. Caricature the continually changing as intervals of steadiness punctuated by instantaneous jumps, then decrease the intervals and the magnitude of the jumps until the phenomenon is smoothed out into a gradual continual change. Recall Galileo's treatment of the continually increasing velocity of free fall. Accordingly we, as Newton, conceive of continually decreasing gravitational pull as the limiting case of intervals of steady pull punctuated by instantaneous decreases of pull.

How are we to compute the effect of an interval of steady pull on our rocket's velocity? What is the key concept here? To this we also find the answer in Galileo: Number 3.1.5, Conservation of Energy. There we have

(7)
$$\tfrac{1}{2}mv^2 = mg \cdot H.$$

$\frac{1}{2}mv^2$, we recall, is the kinetic energy of a mass in moving with velocity v; mg is the force exerted by gravity on this mass at the Earth's surface; H is the height through which the mass falls. When the mass m falls a distance H, its loss in potential energy, $mg \cdot H$, is converted into kinetic energy, $\frac{1}{2}mv^2$. Alternatively put, gravity exerts a force mg on m over a distance of H in the direction of the force; i.e., $mg \cdot H$ is the work done by gravity. And since the free-falling mass starts from rest its initial velocity, and consequently, initial kinetic energy, is zero. So (7) may be read as

gain in kinetic energy = work done in direction of the force.

Alternatively, considering the sequence of events to occur in the reverse

order, so that the mass is thrown up with initial velocity v and work is done in the opposite direction, (7) is to be construed as

loss in kinetic energy = work done against gravity.

Isn't this just what our problem needs for each interval of constant gravitational pull? We must add that Newton was familiar with the concepts of kinetic energy and work done, but he was, of course, unfamiliar with the terminology. Contrariwise, many schoolboys are familiar with the terminology, but not the concepts.

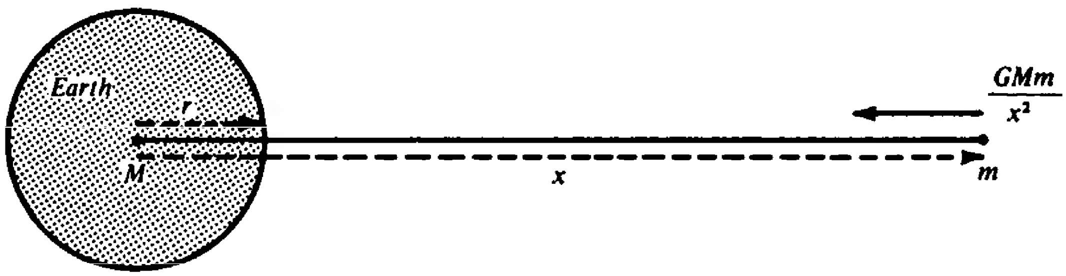

Figure 3.34

Let M be the mass of the Earth, r its radius, and m the mass of our rocket. See Fig. 3.34. We recall that the gravitational attraction of a uniform sphere acts as if the whole of its mass were concentrated at its center, so that the pull F on our rocket at a distance x from the Earth's center is, by Newton's Law of Gravitation (27),

$$F = \frac{GMm}{x^2}.$$

Since we shall need consider a sequence of n positions of our departing rocket (as did Galileo of a free-falling body), it is convenient to denote the n-th position by x_n and the force exerted on our rocket at that point by F_n:

$$F_n = \frac{GMm}{x_n^2}. \tag{33}$$

Also, it is convenient to denote the work done against gravity by our rocket in moving from x_{n-1} to x_n by $W_{n-1,n}$.

Since the initial position of our rocket at blast off is on the Earth's surface, we put $x_0 = r$, and we suppose it to reach a distance R from the Earth's center in the n-th position, $x_n = R$. (We let it voyage first to R, later to infinity.) Our problem is primarily to compute the work done against gravity by our rocket in moving from $x = r$ to $x = R$. We

graph the equation of the force $F = GMm/x^2$ by plotting the points where abscissae are $x_0, x_1, \ldots, x_n$ and complete inner and outer rectangles in relation to these points as shown in Fig. 3.35.

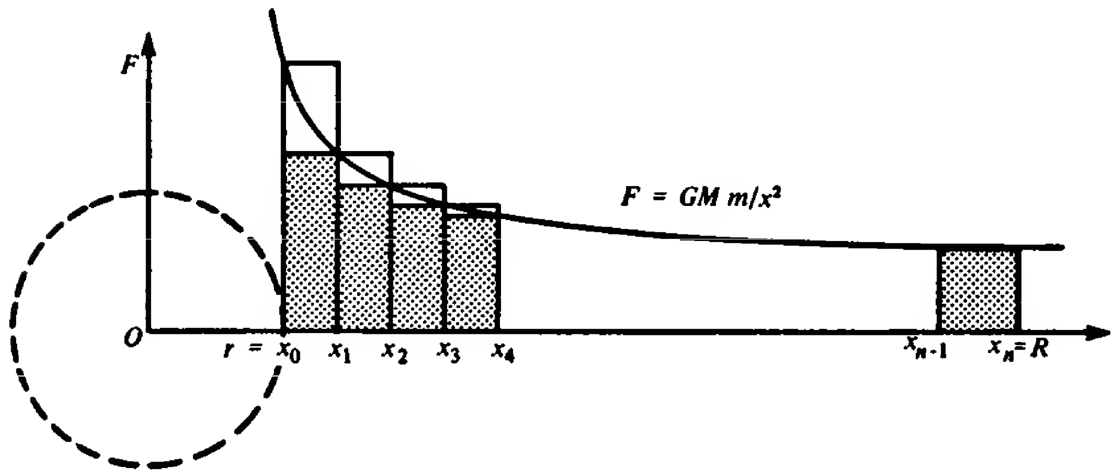

Figure 3.35

What is the work done against gravity by our rocket in moving from x_0 to x_1; what is $W_{0,1}$? The distance from x_0 to x_1 is $(x_1 - x_0)$ and the initial force is F_0. If this remained constant the work done would be $F_0(x_1 - x_0)$, but we know that the actual force continuously decreases. So? It is clear that

$$W_{0,1} < F_0(x_1 - x_0).$$

F continuously decreases from F_0 to F_1; the least force exerted is F_1, and that only for an instant. So? It is clear that

$$F_1(x_1 - x_0) < W_{0,1}.$$

Combining these results, we find that

$$F_1(x_1 - x_0) < W_{0,1} < F_0(x_1 - x_0).$$

Similarly, we estimate $W_{1,2}, W_{2,3}, \ldots, W_{n-1,n}$:

$$\begin{array}{c} F_2(x_2 - x_1) < W_{1,2} < F_1(x_2 - x_1) \\ F_3(x_3 - x_2) < W_{2,3} < F_2(x_3 - x_2) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ F_n(x_n - x_{n-1}) < W_{n-1,n} < F_{n-1}(x_n - x_{n-1}). \end{array}$$

Adding, what do we get?

What is the geometric significance of $F_1(x_1 - x_0)$? This is the area of the rectangle with base $(x_1 - x_0)$ and height GMm/x_1^2, shaded in Fig. 3.35. Similarly, $F_2(x_2 - x_1)$ is the area of the rectangle with base $(x_2 - x_1)$ and height F_2. In short, the sum of the left-hand side elements of our inequalities is the area of the sequence of rectangles just reaching the curve. Let I_n be the area of this inner staircase of n steps. And what is the geometric significance of $F_0(x_1 - x_0)$? This is the area of the rectangle with base $(x_1 - x_0)$ and height F_0. And $F_1(x_2 - x_1)$? Similarly, the sum of the right-hand side elements of our inequalities is the area of the sequence of rectangles just containing the curve $F = GMm/x^2$. Let O_n be the area of this outer staircase of n steps. Adding, we get

$$I_n < W_{0,1} + W_{1,2} + W_{2,3} + \cdots + W_{n-1,n} < O_n,$$

i.e.

$$I_n < W < O_n, \tag{34}$$

where W is the total work done against gravity by our rocket in traveling from x_0 to x_n, that is, in traveling from $x = r$ to $x = R$.

What happens as n, the number of positions considered, is increased? The steps become more numerous and inner and outer staircase more closely coincident. Isn't it clear that for sufficiently large n, the difference between O_n and I_n becomes arbitrarily small? But, it is equally evident that

$$I_n < \text{area under curve} < O_n. \tag{35}$$

So what do you conclude from (34) and (35)? We must conclude that for sufficiently large n, W and the area under the curve are squeezed arbitrarily close; that

$$\text{work done against gravity} = \text{area under curve from } r \text{ to } R.$$

Symbolically,

$$W = [\text{area}]_r^R. \tag{36}$$

The problem to compute W becomes the problem to compute the area under the curve.

Reconsider (34), (35). These are of the form

$$I_n < X < O_n.$$

The method is to find X—and squeeze.

First, to find an X. (34) is a summation of inequalities such as

$$F_1(x_1 - x_0) < W_{0,1} < F_0(x_1 - x_0).$$

Can we, more modestly, find an X_1 such that

$$F_1(x_1 - x_0) < X_1 < F_0(x_1 - x_0)?$$

If so—and if similar quantities $X_2, X_3, \ldots, X_n$ can be similarly found, there will be nothing left to do but add, getting $X = X_1 + X_2 + \cdots + X_n$.

What is F_1? Remember that we are dealing with the curve

$$F = \frac{GMm}{x^2}.$$

(Do not omit m, the constant is not for General Motors.) By (33)

$$F_1 = \frac{GMm}{x_1^2}, \qquad \text{also} \qquad F_0 = \frac{GMm}{x_0^2}.$$

So,

$$F_1(x_1 - x_0) = \frac{GMm}{x_1^2}(x_1 - x_0) \quad \text{and} \quad F_0(x_1 - x_0) = \frac{GMm}{x_0^2}(x_1 - x_0).$$

What quantity lies between them? Not helpful? Well, try a change of emphasis. We rewrite the right-hand sides of these equations:

$$GMm\left(\frac{x_1 - x_0}{x_1^2}\right) \qquad \text{and} \qquad GMm\left(\frac{x_1 - x_0}{x_0^2}\right).$$

Helpful? Try another change of emphasis; concentrate on the similarities (or on the dissimilarities)

$$\left[GMm(x_1 - x_0)\right]\frac{1}{x_1^2} \qquad \text{and} \qquad \left[GMm(x_1 - x_0)\right]\frac{1}{x_0^2}.$$

What lies between them? Obviously something of the pattern

$$\left[GMm(x_1 - x_0)\right]\frac{1}{Y}.$$

The immediate problem is to find a Y such that

$$\frac{1}{x_1^2} < \frac{1}{Y} < \frac{1}{x_0^2}.$$

At this stage one needs luck and a keen sense of what is appropriate, of the fitness of things. Isn't the following a more apt formulation? Find a y such that

$$\frac{1}{x_1^2} < \frac{1}{y^2} < \frac{1}{x_0^2}.$$

Suppose that there is such a y; then

$$\frac{1}{x_1^2} < \frac{1}{y^2} \quad \text{and} \quad \frac{1}{y^2} < \frac{1}{x_0^2},$$

so that

$$y^2 < x_1^2 \quad \text{and} \quad x_0^2 < y^2,$$

and consequently, since we are dealing with positive quantities,

$$y < x_1 \quad \text{and} \quad x_0 < y.$$

I.e.

(37) $$x_0 < y < x_1.$$

These steps are reversible, so that any y satisfying the latter condition also meets all the former conditions.

Can we find such a y? We do know

$$x_0 < x_1,$$

and the abundant occurrence of squares does (whether helpful or not) suggest introducing x_0^2 and x_1^2. Multiplying this inequality first by x_0, and secondly by x_1, we have

$$x_0x_0 < x_0x_1 \quad \text{and} \quad x_0x_1 < x_1x_1;$$

i.e.,

$$x_0^2 < x_0x_1 < x_1^2,$$

so that

$$x_0 < \sqrt{x_0 x_1} < x_1.$$

Thus $y = \sqrt{x_0 x_1}$ or $y^2 = x_0 x_1$ meets our requirement, (37).

In short, since $x_0 < x_1$, $x_0 x_1$ is such that

$$\frac{1}{x_1^2} < \frac{1}{x_0 x_1} < \frac{1}{x_0^2},$$

and

$$\frac{GMm(x_1 - x_0)}{x_1^2} < \frac{GMm(x_1 - x_0)}{x_0 x_1} < \frac{GMm(x_1 - x_0)}{x_0^2},$$

i.e.

$$F_1(x_1 - x_0) < GMm\left(\frac{x_1}{x_0 x_1} - \frac{x_0}{x_0 x_1}\right) < F_0(x_1 - x_0),$$

that is

$$F_1(x_1 - x_0) < GMm\left(\frac{1}{x_0} - \frac{1}{x_1}\right) < F_0(x_1 - x_0).$$

Of course we can deal similarly with the subsequent intervals. The pattern is obvious. All told, we have,

$$\begin{aligned}
F_1(x_1 - x_0) &< GMm\left(\frac{1}{x_0} - \frac{1}{x_1}\right) < F_0(x_1 - x_0)\\
F_2(x_2 - x_1) &< GMm\left(\frac{1}{x_1} - \frac{1}{x_2}\right) < F_1(x_2 - x_1)\\
F_3(x_3 - x_2) &< GMm\left(\frac{1}{x_2} - \frac{1}{x_3}\right) < F_2(x_3 - x_2)\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
F_n(x_n - x_{n-1}) &< GMm\left(\frac{1}{x_{n-1}} - \frac{1}{x_n}\right) < F_{n-1}(x_n - x_{n-1}).
\end{aligned}$$

Adding, we obtain

$$I_n < GMm\left(\frac{1}{x_0} - \frac{1}{x_n}\right) < O_n,$$

i.e.,

$$I_n < GMm\left(\frac{1}{r} - \frac{1}{R}\right) < O_n. \tag{38}$$

Note how neatly all the x's other than the first of the first difference and the last of the last difference canceled out. We are luckier than we knew.

Equation (38) gives our X; to complete the method it remains to squeeze I_n and O_n together. Since for sufficiently large n, the difference between the area of the inner, contained and the outer, containing staircase of the curve becomes arbitrarily small, by (38) we have

$$GMm\left(\frac{1}{r} - \frac{1}{R}\right) = [\text{area}]_r^R.$$

By (36),

$$W = GMm\left(\frac{1}{r} - \frac{1}{R}\right).$$

And since

$$\text{loss in kinetic energy} = \text{work done against gravity}$$

$$\tfrac{1}{2}mv^2 = W,$$

giving

$$\tfrac{1}{2}mv^2 = GMm\left(\frac{1}{r} - \frac{1}{R}\right)$$

or

$$v = \sqrt{2GM\left(\frac{1}{r} - \frac{1}{R}\right)},$$

where v is the initial velocity for our rocket to reach a point in outer space distance R from the Earth's center before being pulled back. For complete escape R must be infinitely large and $1/R$, zero. So the escape velocity $\bar{v}$ is given by

$$\bar{v} = \sqrt{\frac{2GM}{r}}. \tag{39}$$

We have solved our problem. In solving it we circumvented the technique of, but not the ideas basic to, integration. We did it in the

old-fashioned way which really goes right back to Archimedes. To compute the required area we really needed tremendous luck; with integral calculus such problems become merely routine—and less exciting. I hope you will use this method to pave the way for the integral calculus. Concentrate on the essential concepts: area under a curve, successively better approximations, and the specific, lucky inequaliy.

3.4.7 Ratio of Escape and Orbital Velocities

Which is greater, the go-away or the go-around velocity? At the beginning of this section you were asked to commit yourself to an opinion. We are now in a position to determine if you are right.

The roof-top orbital velocity v is given by (21),

$$v = \sqrt{g_E \cdot r}\,.$$

And m being the mass of our rocket and M the mass of the Earth, by Newton's Law of Gravity (27), we have

$$m \cdot g_E = \frac{GMm}{r^2}\,,$$

that is,

$$g_E = \frac{GM}{r^2}\,.$$

Substituting in (21), we get

$$v = \sqrt{\frac{GM}{r^2} \cdot r} = \sqrt{\frac{GM}{r}}\,.$$

And by (39)

$$\bar{v} = \sqrt{\frac{2GM}{r}} = \sqrt{2} \cdot \sqrt{\frac{GM}{r}}\,,$$

so that

$$\frac{\bar{v}}{v} = \sqrt{2}$$

i.e.,

$$\bar{v} = \sqrt{2} \cdot v.$$

Perhaps it is obvious that the go-away velocity is greater than the go-around velocity, but it is not obvious that it is $\sqrt{2}$ times as great.

We have seen how Newton, primarily by relating the motion of falling apple, cannon ball, and the Moon to Kepler's Laws, was led to his Law of Universal Gravitation and Theory of Dynamics. Are not the Luniks, Sputniks, Pioneers, etc. voyaging overhead a fitting memorial to, as well as a vindication of, his genius?

CHAPTER FOUR

Physical Reasoning in Mathematics

To date, what have we done? First we discussed measurement, especially in astronomy; then simple but pervasive topics culled from the history of statics, and finally, great discoveries from the history of dynamics—so many of which hark back to the stars. We have seen something of the role played by mathematics in the development of science; that the aim of physics is to condense its knowledge into mathematical formulae; that, as Galileo so delightfully expressed it, the book of Nature is written in mathematical characters.

Yet this view, although undeniable, is one-sided—or should I say unidirectional? Of course mathematics helps physics. But you must not suppose that help always flows downstream from mathematics to physics; the river of thought is tidal. My object in this chapter is to navigate an incoming tide, to show how help flows also from physics to mathematics.

My lecture-room navigation will not be reproduced here as my upstream voyage is already carefully charted in my *Mathematics and Plausible Reasoning*, Vol. 1, pp. 142–167, to which the interested mariner is directed.

CHAPTER FIVE

Differential Equations and Their Use in Science

The reader should be somewhat familiar with the concepts and the techniques of integral and differential calculus; yet knowledge of the theory of differential equations is not a prerequisite. What such equations are and how they must be treated will be explained (roughly but sufficiently for our purpose) later, when they naturally emerge from physical problems. It will turn out that differential equations are useful in science. We cannot understand how and why they are useful before we have used them.

SECTION I. FIRST EXAMPLES

5.1.1 Rotating Fluid

One lump of sugar, or two? Cream? We have all observed a lady taking tea. What happens? The faster she stirs, the higher up the side of her cup the tea climbs. If she stirs too fast she spills it and ruins an afternoon. Her teacup contains a problem for her and a problem for us. Our problem is amenable to mathematical treatment: What is the surface shape of the rotating tea?

First consider a motionless liquid. We have all seen a glass of water when no one is kicking the table. Its surface looks flat, yet closer examination shows its surface to be not entirely horizontal; it curls up ever so slightly at the edges, due to surface tension. For water substitute mercury, and surface tension causes precisely the opposite effect, a curling down at the edges. A phenomenon distinctly visible in a mercury barometer. See Figs. 5.1(a) and 5.1(b). The point, made so many times, is

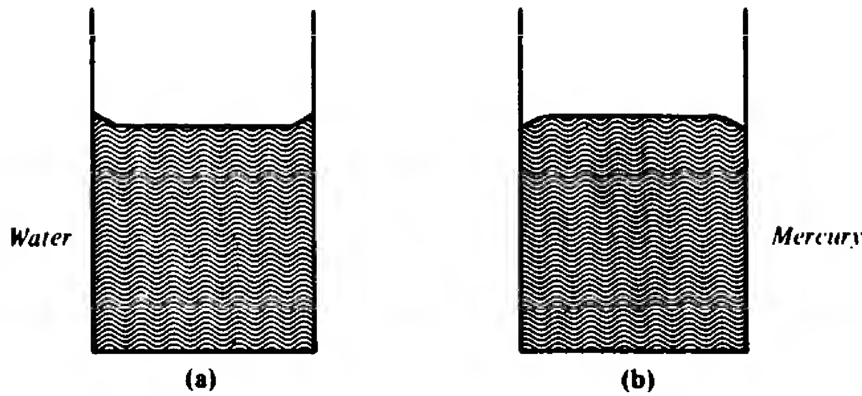

Figure 5.1

that mathematics succeeds in dealing with tangible reality by being conceptual. We cannot cope with the full physical complexity; we must idealize. We neglect the minor circumstance of surface tension, we suppose the surface of a non-rotating fluid to lie wholly in a horizontal plane.

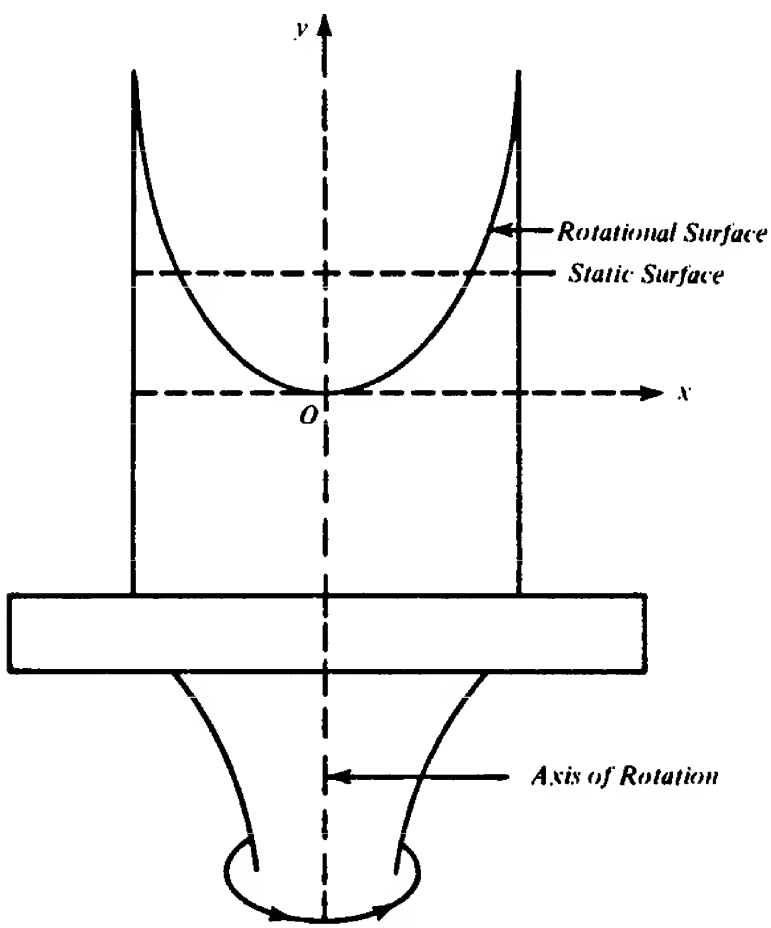

Figure 5.2

Instead of the tea stirred by the lady, let us observe a more clear-cut experiment. We set a glass half full of colored water well centered on a

centrifugal machine, set it in uniform motion about a vertical axis and observe what happens. At the center of the rotating glass the water is depressed, at the circumference, elevated. If the rate of rotation is increased, the water rises higher towards the rim and sinks lower in the middle, forming a deeper hollow. See Fig. 5.2. A cross section (in a vertical plane through the axis of rotation) looks roughly as illustrated by Fig. 5.2. With a brightly colored liquid there is no problem to see what it looks like; the problem is to state precisely what it is, to give a mathematical description of the shape of the rotating surface.

Another crude description is to say that the surface shape is hemispherical—a remark that introduces the notion of surface of revolution. If hemispherical, what would a cross section be? Yes, a semicircle, the same as any other cross section; the surface generated by rotation of a semicircle about its central radius is a hemisphere. Similarly, the rotation of a circle about a diameter generates a sphere. And, a most interesting case, what happens if a circle is rotated about a line which does not intersect it? An anchor ring, wedding ring—or to be technical rather than nautical or matrimonial—a torus, is generated. See Fig. 5.3.

Since hemisphere, sphere, and torus are solids generated by rotation they are said, appropriately, to be solids of revolution and their surfaces to

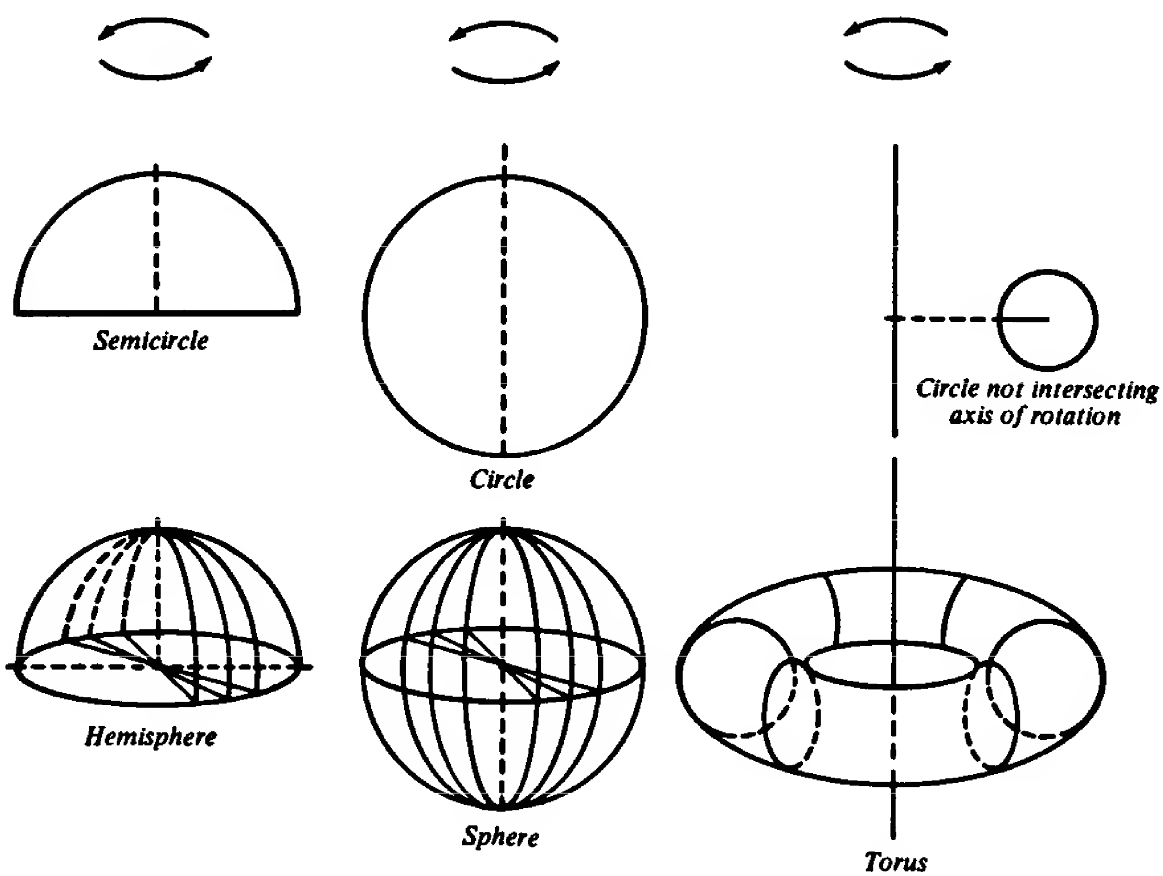

Figure 5.3

be surfaces of revolution. To determine a surface of revolution it is obviously sufficient to know the shape of the plane curve which is rotated and its position relative to the axis of rotation.

Reconsider the rotating surface of our liquid. As the different parts of the surface rotate past our fixed gaze there is no discernible change in shape, nor would the Principle of Insufficient Reason suggest one. So? The rotating surface may be conceived as a surface of revolution, as generated by the curved boundary of any cross section through the axis of rotation. Other cross sections do not trace out new solids, they retrace the same solid. Thus, to determine the surface of the rotating liquid it is sufficient to determine the geometry of any cross section (in a vertical plane through the axis of rotation).

Technically, the surface-generating curved boundary of the solid-generating cross section is said to be a *meridian* of the surface. This term, when fully understood, is helpful as well as technical. Its helpfulness is revealed by its origin. The Latin *meridies* (medius, middle + dies, day), which became the French *midi*, means *noon*. The people who have noon at the same time live on a noon curve or meridian. San Franciscans and Seattleites—or whatever they call themselves—have the Sun at its zenith, twelve o'clock noon, solar time, at the same time. New Yorkers have the Sun highest in the sky some three hours earlier; they live on a different meridian. A meridian is any great circle connecting North and South Poles, a curve that, if rotated, generates the Earth's spherical surface.

The upshot: our problem becomes that of finding the equation of the meridian. How are we to do this? Yes, we begin by introducing a rectangular coordinate system. And where are we to put the axes of coordinates? It seems natural to make the y-axis the (vertical) axis of rotation. Consequently, the x-axis will be horizontal. And what about the origin O? There remains some choice. Foreseeing that the axis of rotation goes through the center of the hollow, it would appear convenient to make O this central point, see Fig. 5.2. Yet observe that although we are free to put the y-axis anywhere we please and are free to put O anywhere on the y-axis we please, the two freedoms are not, so to speak, equally free. A not unnatural alternative choice for O is the point of intersection of the axis of rotation with the original horizontal surface of the liquid. But what is an alternative for the y-axis? True, we could make it the edge of our container, but is this a genuine alternative? Does it really please us? Although logically possible, it is physically abhorrent. The reader who finds this point farfetched lacks feeling for what is of physical importance. We insert rectangular axes as illustrated in Fig. 5.2.

Our problem has become even more specific: to find the equation of the meridian relative to these axes of coordinates. While this is our

immediate interest, let us not lose sight of our greater ambition. Remember that these lectures are entitled "Mathematical Methods in Science"; our major aim is to learn how to apply mathematics to physics.

Behind the special problem before us we wish to see something more general–the typical attitude of the scientist who uses mathematics to understand the world around us.

In the solution of a problem such as that of the rotating fluid there are typically three phases. The first phase is entirely or almost entirely a matter of physics; the third, a matter of mathematics; and the intermediate phase, a transition from physics to mathematics. The first phase is the formulation of a physical hypothesis or conjecture; the second, its translation into equations; the third, the solution of the equations. Each phase calls for a different kind of work and demands a different attitude. To exemplify and to amplify these remarks we continue with our problem.

The Physical Phase Mathematical textbooks intend to propose well-formulated problems. If the aim of such a problem is the determination of an unknown, the textbook should state unambiguously the data and the condition by which the unknown is determined.

We have now, however, a physical problem. We want to find the equation of the meridian curve whose rotation generates the curved free surface of the revolving fluid. This equation is our aim, our unknown, that much is clear. Yet what is given? What is the condition? We observe the rotating fluid, yet the phenomenon observed does not give us anything by itself. We have to take what we need. Especially, we have to produce the concepts, the terms in which the condition determining the unknown must be expressed.

I do not suppose here that the reader knows hydrodynamics. On the contrary, if he knows it, he should pretend that he has forgotten it. Facing any part of the observable reality, we are never in possession of complete knowledge, nor in a state of complete ignorance, although usually much closer to the latter state. If we deal with our problem not knowing, or pretending not to know, the general theory encompassing the concrete case before us, if we tackle the problem "with bare hands", we have a better chance to understand the scientist's attitude in general, and especially the task of the applied mathematician.

I shall assume, however, that the reader is somewhat familiar with the elementary dynamics of a material point.

If you cannot solve the proposed problem, try to solve first a simpler, related problem. Why does the free surface of the rotating fluid form a round hole? Well, why does the free surface form a piece of a horizontal plane when the fluid is not rotating but is at rest? (We neglect the surface

tension.) Consider a bit of the fluid at rest, a minute portion of it adjoining the horizontal free surface. It is at rest, it does not move, although its weight, the gravitational force, acts on it, urging it vertically downward. What is counteracting the weight? It must be a force pointing vertically upward, perpendicular to the horizontal free surface, and it cannot be but the resultant, say R, of the pressures exerted by the contiguous parts of the fluid. We conclude that the *resultant pressure* R *is perpendicular to the free surface of the fluid.*

Now let us return to our problem of the rotating fluid. We consider a bit of the rotating fluid, a minute portion of it adjoining its free surface at a certain point P. This bit of the fluid rotates uniformly, describes a horizontal circle with constant angular velocity. We know from dynamics that such a uniform circular motion supposes that the rotating mass is acted on by a centripetal force of constant magnitude directed to the center of the circular path. How does such a centripetal force arise?

It is natural to assume that the *centripetal force is the resultant* of those *two forces* we already know to be acting at the bit of fluid considered: *Gravity*, pointing vertically downward, and the *resulting pressure* R, perpendicular to the free surface, due to the action of the contiguous parts of the fluid. See Fig. 5.4.

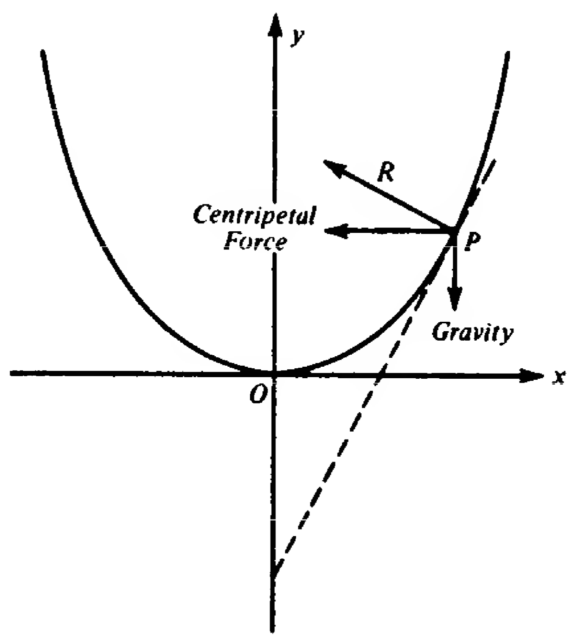

Figure 5.4

We have arrived so at an essential physical idea that could explain the phenomenon. Yes, it *could* be an explanation—but is it one really? The question is justified, but let us postpone it. Let us first restate our explanation more fully, more precisely, in quantitative terms. We have now completed the first, physical, phase of our inquiry and must begin

the next phase which should bring us the mathematical formulation of the physical idea conceived.

The Transition from Physics to Mathematics We have to restate the physical idea conceived in the foregoing more fully and more precisely, so that we can eventually express it in mathematical symbols.

Let m be the mass and v the uniform circular velocity of the bit of fluid or particle P at the point (x, y) on the meridian. What is the distance of P from the axis of rotation? Yes, its abscissa x. Consequently, since its circular motion is uniform, its centripetal acceleration has magnitude v^2/x. And since

$$\text{force} = \text{mass} \times \text{acceleration}$$

the centripetal force exerted on m has magnitude mv^2/x, and the gravitational force mg, where g, as usual, stands for the gravitational acceleration. Thus, in addition to the directions of all three forces acting at P being depicted by Fig. 5.4, we know the magnitudes of two of them. Moreover, the centripetal force is the resultant of the other two. This is the decisive idea discovered in the foregoing phase. So, to depict the relations between these three forces we need merely complete the parallelogram of which a horizontal line from P (which represents the centripetal force) is a diagonal and a vertical line from P (which represents the gravitational force) is a side. It is sufficient to remark that PC must be parallel and equal to AB. We have Fig. 5.5.

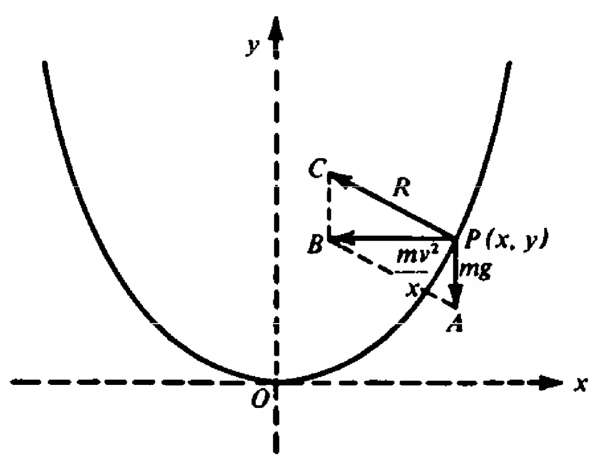

Figure 5.5

What next? Our physical hypothesis is that the shape of the meridian is conditioned by the situation depicted in Fig. 5.5. How are we to deduce the equation of the meridian from the depicted circumstances of P? It is at least clear that we must make use of the geometry of this figure and take into account the coordinates of P. But, of the relevant items

(mv^2/x and mg) only one contains x and neither contains y. Perhaps we should console ourselves with the thought that half a loaf is better than no bread. Whatever our disconsolations, we must utilize the item mv^2/x, for otherwise we have no prospect of incorporating x in an equation. Yet on second thought this is not a happy prospect, for with x comes m. It would be most embarrassing if the shape of the curve should prove to depend upon the mass of the particles along it. Embarrassing because mass will depend upon size—and our supposition is that the dimensions of the bits of fluid at P are negligible. So? While hanging on to x we must get rid of m.

How are we to eliminate m? Take another look at Fig. 5.5. Consider $\triangle ABP$. The ratio of mv^2/x to mg is independent of m. Doesn't this suggest that the tangent or cotangent of angle A ought to be of interest to us? Or, the opposite angles of a parallelogram being equal, angle C? Tangent of an angle? Wait a minute. Why, of course. We have neglected an important fact discovered in the physical phase: that R is *perpendicular to the meridian*, that R is perpendicular to the *tangent* at P.

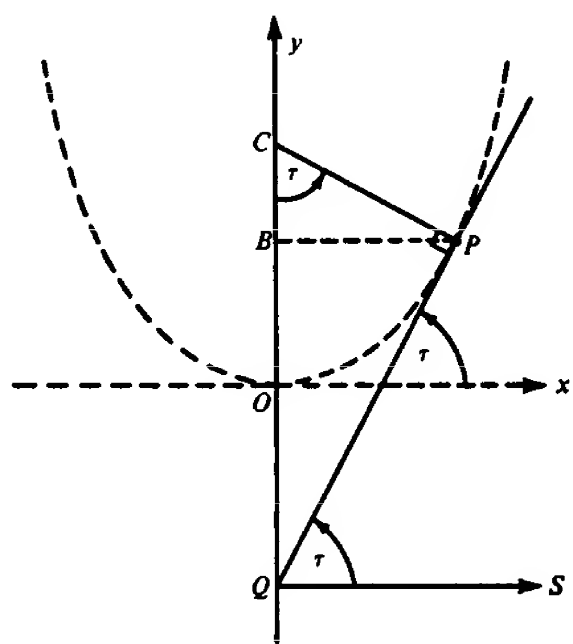

Figure 5.6

Let us introduce this tangent shown in Fig. 5.4. At the same time, since we are at liberty to select any scale we please for the parallelogram of forces, it is convenient although not essential to have B (and consequently C) on the y-axis. We consider Fig. 5.6. Let the tangent to the meridian at P cut the x-axis at an angle τ (τ is the Greek T and T is for *Tangent*) and meet the y-axis at Q, and let QS be parallel to Ox. It follows that $\angle SQP = \tau$. Since angles CPQ and CQS are both right angles, angles QCP and PQS are both complements of $\angle CQP$. It

follows that $\angle C = \tau$. Hence

$$\tan \tau = \tan C = \frac{mv^2/x}{mg} = \frac{v^2}{xg}$$

so that

$$\tan \tau = \frac{v^2}{xg}. \tag{1}$$

Partial success. We have an equation involving x.

Obviously we would like to substitute a function of x and y for $\tan \tau$. Can we? Just about the first thing we learn in the differential calculus is that the first derivative dy/dx is the slope of the tangent to the curve $y = f(x)$ at the point (x, y). And since τ is the angle the tangent line at P makes with the x-axis, $\tan \tau$ is the slope of this tangent line. Consequently,

$$\frac{dy}{dx} = \tan \tau.$$

Hence, by (1),

$$\frac{dy}{dx} = \frac{v^2}{xg}. \tag{2}$$

The left-hand side of our equation now contains x and y, even if not in the straightforward way to which we are accustomed in dealing with algebraic formulae. Perhaps we should take comfort in the reflection that an unaccustomed appearance is preferable to no appearance at all.

Other aspects of (2) also provide food for thought. Let us suppose ourselves to be looking down upon the rotating fluid from way up above the y-axis. What do we see? Conceptually, we view the surface of revolution as hosts of particles rotating in circles concentric about O (for the axis, from directly above, appears as a point). All the particles in the same horizontal plane revolve with the same velocity, but the higher the plane, the greater the radius of rotation x and the greater the velocity v. In other words, v is a function of x. But what is the really remarkable feature? That any two particles on a meridian continue to revolve with the meridian. They behave as if they were the tips of clock hands that turn to keep pace with one another. See Fig. 5.7. While one particle circles about O from P_1 to P_1', the other circles about O from P_2 to P_2'. The two particles rotate with the same angular velocity.

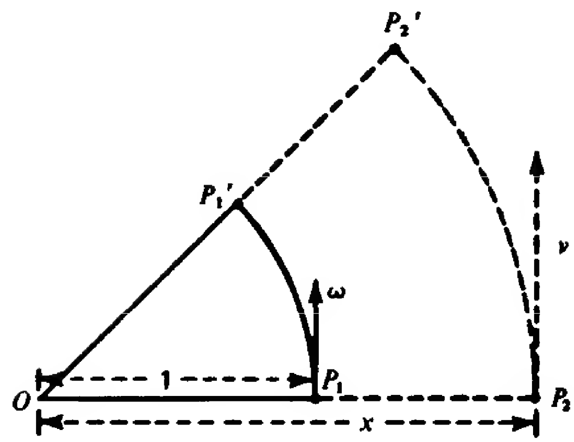

Figure 5.7

And how do we measure angular velocity? A particle is said to have angular velocity ω if it traverses an arc subtended by ω radians of a unit circle in unit time. We use the Greek letter ω (Omega), as is customary in these matters. The reader may reflect that since the accuracy of a clock depends upon the rates of rotation of its hands and not on their lengths, Omega would be an appropriate name for a company manufacturing chronometers.

Let us suppose that $OP_1 = 1$, $OP_2 = x$, and that one particle travels along arc P_1P_1' in unit time with angular velocity ω while another travels along arc P_2P_2' with velocity v. It follows that

$$\text{arc } P_1P_1' = \omega, \qquad \text{arc } P_2P_2' = v.$$

But obvious to intuition—and by a theorem of Euclid—these arc lengths are proportional to their radii. Thus

$$\frac{\omega}{1} = \frac{v}{x}$$

so that

$$v = \omega x. \tag{3}$$

It remains to remark that since the clock hands are synchronous both particles have the same angular velocity, so that the particle with velocity v and rotational radius x also has angular velocity ω.

Squaring (3) and substituting in (2), we have that $dy/dx = \omega^2x^2/xg$, i.e.,

$$\frac{dy}{dx} = \frac{\omega^2 x}{g}. \tag{4}$$

And what are the implications of this equation? With ω constant (g being constant, of course), dy/dx increases or decreases as x increases or decreases; for any steady rate of rotation the surface of the fluid near the y-axis is flatter, that farther away, steeper. And note that for a given x, if ω is increased, dy/dx is increased. Thus (4) implies that the faster the lady stirs her tea, the steeper the sides of the hollow become. If she does not stir, $\omega = 0$, the slope is horizontal and the surface flat. These implications are in accordance with the obvious facts; they afford grounds for accepting (4) as a plausible mathematical statement of the condition upon which the shape of the meridian depends. But according to (4) the shape of the meridian is also dependent upon g. It implies that if g were reduced to one-sixth its terrestial value the meridian would become six times as steep. Although we do not yet take afternoon tea on the Moon, and an appreciable change in g is outside our common experience, we have reason to suspect that tea is more readily spilt at Lunar tea parties. Our grounds are substantial, not conclusive. We are disposed to accept (4) as correct.

Furthermore, we can apply the dimensions test. It is helpful to think of the various entities in terms of units in which they could be measured. We take centimeter and second as the units of length and time; mass is not involved. And do remember that angular velocity is measured in radians per unit time. Since the radian measure of a central angle is the length of the intercepted arc divided by the radius of the circle, it (like slope) is the ratio of two lengths; its dimensions are zero.

$$\frac{dy}{dx} = \tan \tau = \text{slope} = \frac{\text{cm}}{\text{cm}} = \frac{L}{L} = L^0 = 1,$$

so that the left-hand side of (4) has zero dimensions.

$$x = \text{cm} = L,$$

$$\omega = \frac{\text{radian}}{\text{sec}} = \frac{1}{T} = T^{-1}, \qquad \omega^2 = T^{-2},$$

$$g = \text{acceleration} = \frac{\text{cm}}{\text{sec}^2} = \frac{L}{T^2} = LT^{-2},$$

so that

$$x\omega^2 \frac{1}{g} = LT^{-2} \frac{1}{LT^{-2}} = 1.$$

It checks. We accept (4) as correct.

Our final mathematical statement, spelled out in full, is that the generating meridian of a fluid that rotates with angular velocity ω in a gravitational field g is such that any point (x, y) of the meridian satisfies the condition (4)

$$\frac{dy}{dx} = \frac{\omega^2 x}{g}.$$

This completes the second phase.

Oh yes, we have been a long time arriving at this statement. But we were not told to use the notion of meridian; we were not told to think of a liquid as a conglomeration of point mass particles; we were not given the angular velocity; we were not given the gravitational field. All these things we had to take for ourselves. The problem was to decide which ingredients to take. If an investigator is clear about their relevance at the outset, his work is routine; he does not have a problem.

The Mathematical Phase The final phase is essentially mathematical: to deduce the equation of the meridian curve from (4)

$$\frac{dy}{dx} = \frac{\omega^2 x}{g}.$$

The novelty of this equation as opposed to our familiar algebraic equations is that it contains the differential coefficient dy/dx. For this reason mathematicians term it a *differential equation*. We have reached the principal topic of this chapter.

The reader may anticipate that since dy/dx may alternatively be termed first derivative, (4) may alternatively be termed *derivative equation*. This is never done. The latter terminology would invite confusion for all equations are in a sense derivative—from the given, or, as here, the taken. The former is firmly indicative that the differential calculus is involved. But, with equal propriety,

$$\frac{d^2y}{dx^2} = \frac{y}{x^2}$$

is, for example, to be termed a differential equation. This equation contains a second derivative whereas (4) contains no derivative of higher order than the first. It turns out that the order of the highest derivative involved has an important bearing on the solution of the equation. Accordingly, distinction is made: (4) is said to be a first-order differential equation, the succeeding example, second order.

Differential equations of the first order are indicated schematically by

$$F\left(x, y, \frac{dy}{dx}\right) = 0. \tag{5}$$

This exhibits that the ingredients are the independent variable x, the dependent variable y, and the first derivative of y with respect to x. The F is to emphasize that there is a functional relation between these ingredients. Note that from (4), we have

$$\frac{dy}{dx} - \frac{\omega^2 x}{g} = 0,$$

an especially simple case of (5), for we have no ingredient y. Unfortunately there is no general method of solution of differential equations of all orders, nor even a general method for just those of the first order. It is convenient to classify the latter according to the methods to which their solutions are amenable. (4) is readily solved by the method of *separation of the variables*.* (This explains why I chose to introduce the topic of differential equations via the problem of a rotating fluid.)

Separation of the Variables The first derivative, the result of the differentiation of y with respect to x, was written by Leibniz in the form

$$\frac{dy}{dx}$$

(other notations are y', $\dot{y}$, Dy, $D_x y$). Leibniz's notation deserves some comments, because it is both extremely useful and dangerous.

Today, as the concepts of limit and derivative are sufficiently clarified, the use of the notation dy/dx need not be dangerous. Yet, the situation was different in the 150 years between the discovery of the calculus by Newton and Leibniz and the time of Cauchy. The derivative dy/dx was considered as the ratio of two "infinitely small quantities", of the "infinitesimals" dy and dx. Such consideration was helpful: it greatly facilitated the systematization of the rules of the calculus and gave intuitive meaning to its formulas. Yet this consideration was also obscure—so obscure and nebulous, in fact, that it brought mathematics into disrepute: some of the best minds of the age, such as the philosopher Berkeley, complained that the calculus is incomprehensible.

*"Not a method, just a trick" says a voice which tries to sound superior. See the footnote at the end of subsection 5.1.4, p. 192.

It should be made clear to any beginner today that dy/dx is the *limit* of a ratio and emphatically *not* the ratio of dy to dx: the full symbol dy/dx has a clearly defined meaning, but (at least in the first semesters) the best is to consider its parts dy and dx as devoid of meaning: the word WORD has a meaning, but its parts WO and RD have none.

Once we have realized this sufficiently clearly, we may, under certain circumstances, treat dy/dx so *as if it were* a ratio: adults and experts may do things that children or beginners should not do. For instance, we may conveniently recollect the geometric meaning of dy/dx as slope of the tangent to the curve in considering the "infinitesimal" right triangle with horizontal leg dx and vertical leg dy. See Fig. 5.8. We may do so *if* we take such consideration just as a colloquial but short (although somewhat sloppy) expression for a limiting process which we have once carefully considered and could reproduce if needed.

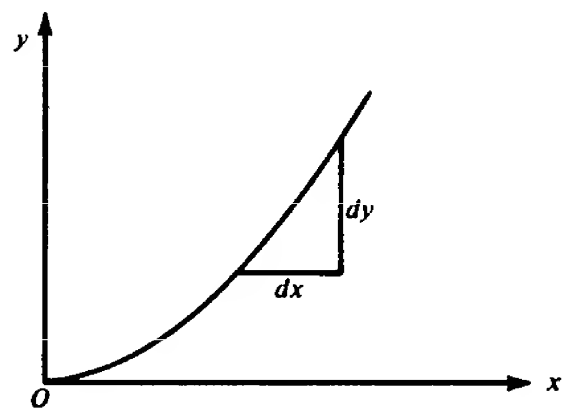

Figure 5.8

Trusting the wisdom of Leibniz's notation, we treat dy/dx as if it were a ratio and multiply (4) by dx to achieve separation of the variables. We get

$$dy = \frac{\omega^2}{g} \cdot x dx.$$

The left-hand side does not now contain x, nor the right y; we can integrate immediately:

$$\int dy = \int \frac{\omega^2}{g} \cdot x dx.$$

And since a constant factor is not affected by integration

$$\int dy = \frac{\omega^2}{g} \cdot \int x dx.$$

And what do we differentiate to get x? Yes, $\frac{1}{2}x^2$. So, not forgetting the constant of indefinite integration, we have

$$y = \frac{\omega^2}{g} \cdot \frac{1}{2} x^2 + C$$

or

$$y = \frac{\omega^2}{2g} \cdot x^2 + C. \tag{8}$$

By assigning different numerical values to C we obtain different curves. Which curve is the required meridian? Recall that we selected our axes such that the origin O is at the bottom of the hollow, i.e., such that $y = 0$ when $x = 0$. This is termed an *initial* or a *boundary* condition. These terms are used because the condition determines the position of the point at which the meridian is initiated or by which the meridian is bounded.

Although our problem is not determined by the differential equation alone, establishing this equation is the major, more responsible work. To obtain it we had to probe a complex physical situation. To the contrary, the initial condition is obvious and somewhat arbitrary. Equation (8) bears testimony that the horizontal plane in which we select our x-axis is a matter merely of notational convenience. As remarked much earlier, our choice, unlike that of the y-axis being the axis of rotation, has no physical significance.

Applying the initial condition, $y = 0$ when $x = 0$, to (8), we obtain

$$0 = \frac{\omega^2}{2g} \cdot 0 + C,$$

so that

$$C = 0,$$

and

$$y = \frac{\omega^2}{2g} \cdot x^2.$$

The meridian is a parabola; the surface, a paraboloid of revolution. We have solved our problem.

5.1.2 Galileo: Free Fall

Our second illustration of the use of differential equations is conveniently provided by Galileo's problem of free fall.

We introduce a vertical x-axis whose positive direction is downwards and suppose a heavy particle to be let fall from the origin O, $x = 0$. As

we let the particle fall we start our stop watch, $t = 0$. Thus the motion is subject to the initial condition that $v = 0$ and $x = 0$ when $t = 0$. How far, t seconds later, will the falling particle be below O? Obviously it can be only in one place at a time, and so long as it continues to fall, it will be in different places at different times. Where depends on when; x is a function of t, $x = f(t)$. Our problem is to specify the function, $f(t)$. See Fig. 5.9.

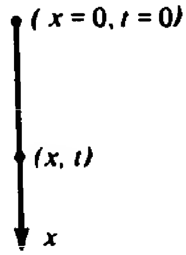

Figure 5.9

Aristotle, among others, noted that the farther a body falls, the faster it falls. Galileo, we recall, was insistent upon being more specific and first made the very natural conjecture that the acquired velocity v is proportional to x, the distance fallen. That is, that

$$v = cx, \tag{9}$$

where c is a constant independent of x. Earlier, in Number 3.1.3, we remarked that Galileo eventually came to the conclusion that this conjecture is not merely wrong but logically absurd. But as he had no calculus he was obliged to invent a highly subtle, not easily accessible argument. Here is an opportunity to present the calculus version.

Since the instantaneous velocity v is given by

$$v = \frac{dx}{dt}, \tag{10}$$

substituting (10) in (9), we obtain the first-order differential equation

$$\frac{dx}{dt} = cx.$$

Again, trusting Leibniz's wisdom, we treat the derivative as a ratio. Multiplying by dt, we have

$$dx = cx \cdot dt.$$

But birds of a feather should flock together. We divide by x.

$$\frac{dx}{x} = c \cdot dt.$$

The variables are now separated. It remains to integrate.

$$\int \frac{1}{x}\, dx = c \int dt.$$

And what do we differentiate to obtain $1/x$? Yes, $\log_e x$. So

$$\log_e x = ct + k,$$

where k is the arbitrary constant of indefinite integration. But we want a formula for x, not $\log_e x$. It is, at this stage, convenient to remember that by definition of logarithm the equations

$$\log_{10} 2 = 0.30103, \qquad 2 = 10^{0.30103}$$

are equivalent. With x instead of 2, e instead of 10, and . . . . Yes, you can do it for yourselves. We get

$$x = e^{ct+k}.$$

But, by the initial conditions, when $t = 0$, $x = 0$, so

$$0 = e^{c0+k} = e^k.$$

If $k > 0$, e^k is clearly positive. If $k < 0$, say equal to $-k'$ where k' is positive,

$$e^k = e^{-k'} = \frac{1}{e^{k'}} = \frac{1}{\text{positive}} = \text{positive}.$$

Thus e^k is necessarily positive. In short: if the free-fall velocity is proportional to the displacement, then

$$0 = \text{a positive number}.$$

This, as Euclid would say, is absurd. Therefore free-fall velocity cannot be proportional to displacement.

Again we have a problem resolved by a differential equation with an initial condition. Despite being effected by such a simple separation of the variables it is historically important. We are apt to suppose that all

untenable physical theories are eventually refuted by experiment. This one was defeated by logic: we have proved that it is inconsistent.

Galileo's second thoughts conjectured that the velocity acquired is proportional to the time taken to acquire it. That is, that

$$v = gt \tag{11}$$

where g is a constant independent of t.

Now substituting (10) in (11) instead of (9), we have

$$\frac{dx}{dt} = gt.$$

We separate the variables by multiplying by dt:

$$dx = gt \cdot dt.$$

It remains to integrate:

$$\int dx = g\int t \cdot dt.$$

We get

$$x = \frac{1}{2}gt^2 + C.$$

And since initially $x = 0$ when $t = 0$

$$0 = \frac{1}{2}g \cdot 0 + C,$$

so that

$$C = 0,$$

and

$$x = \frac{1}{2}gt^2.$$

This is a useful proposition of physics. Reflection upon the contrast between its derivation here and Galileo's (cf. Number 3.1.3) is rewarding. By effecting solutions without having to think what we are really doing we gain a lot—and can lose a lot.

5.1.3 Catenary

Catenary is derived from the Latin *catena* meaning chain, and is used as a technical term for the curve formed by a uniform chain hanging freely from two points not in the same vertical line. Our problem is to specify the shape of the catenary: to determine its equation. With one point of support vertically below the other there is no catenary and no problem; the shape is obvious.

When I was young the well-situated gentleman was wont to indicate his opulence—not to mention emphasis of his corpulence—by sporting a golden watch chain across a wide expanse of waistcoat. But even if golden chains and waistcoats were still with us this would not be sufficient reason to consider the catenary. In this technological age, telegraph wires, and high-tension cables display some important catenaries. Unless the strength of steel has recently been increased it is still the case that a steel cable catenary six miles long would break under its own weight.

Consider the catenary pictured in Fig. 5.10. Obviously the supports at A and A' do much more than take the weight of the cable. Recall the exertion needed to take up some of the "slack" of a sagging clothes line. Many engineers spend much of their working lives calculating tensions in wires and cables, particularly the magnitude and direction of the pulls at the points of support. The effect of contact with a falling high tension cable is apt to be permanent as well as instantaneous.

Not all freely suspended wires hang in catenaries. The word chain in the above definition is used to imply strength and flexibility. We think of a chain not rusted at the links. Ideally it does not stretch and is free to swivel at its linkages; it is inextensible and offers no resistance to bending. And the word uniform implies that it is made of homogeneous material, that its weight per unit length is the same throughout. The more nearly a suspended wire, cable, or chain is flexible, inextensible, and homogeneous, the closer its shape approximates a catenary.

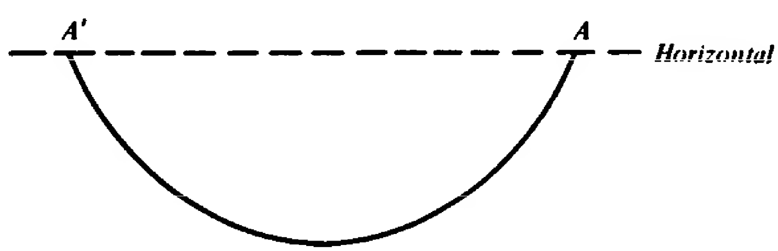

Figure 5.10

We consider a perfect chain suspended from two supports in the same horizontal plane. (The case where the supports are not in the same horizontal plane will be considered subsequently.) Does it hang

lopsidedly? If so, to which support is the bottom of the curve nearer? The left or the right? Yes, we have encountered the Principle of Sufficient Reason before. It will hang symmetrically with respect to the vertical equidistant between its supports. It is natural to take this vertical as the y-axis. And the x-axis? As for a rotating fluid it is tempting although not essential to put it at the bottom of the curve. I prefer to put it an arbitrary distance below; my reason will become apparent later. We have the situation depicted by Fig. 5.11. Our problem is to determine the equation of the catenary, $y = f(x)$, relative to our selected rectangular coordinate axes.

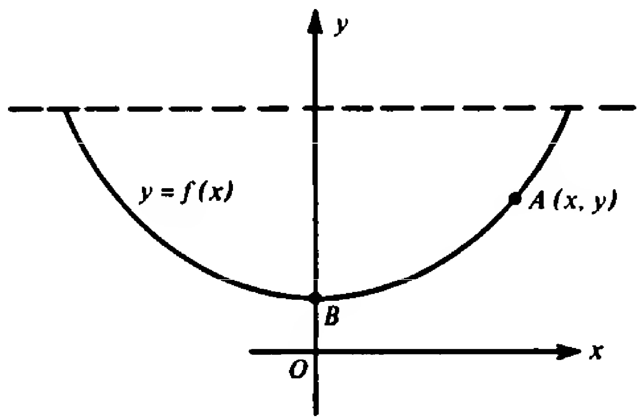

Figure 5.11

The first physical phase is difficult. Although it is clear what we are to find, what we are to take as given is far from evident. Rereading the definition of *catenary* we note that the chain is uniform. We take this to imply that it is made of homogeneous material, so that it has constant weight per unit volume. Yet on second thoughts, *uniform chain* implies more than this: a chain with hefty links at one end and slender ones at the other, or a cable thick at one end and thin at the other, would not be described as *uniform* even if made of homogeneous material. Additionally, it implies uniform cross section: every part of the chain has the same weight per unit length. Idealized, with cross section shrinking until the chain becomes a line, this amounts to saying that it has constant linear density. As is customary we take this density to be λ. (λ is the Greek L and L stands for *Length*.)

Is this sufficient, or do we need provide ourselves with additional data? If a chain has linear density λ, it has it no matter what its shape. It would still have this density while being cracked like a whip. But it is not being cracked like a whip; it isn't moving at all: it is in equilibrium. We conjecture that the condition that a chain has density λ and is in equilibrium is sufficient to determine its shape. Or, not to do violence to the English language, we may reformulate our problem: Given that a

perfect chain of density λ, suspended as shown in Fig. 5.11, is in equilibrium, find its shape. It is of course understood that the answer is to be given relative to our selected coordinate system.

The second, transitional phase is the translation of our conjectured condition, equilibrium, into mathematics. We anticipate ending up with a differential equation with an initial condition.

Unwind some cotton from a cotton reel and pull. The unwound cotton is tangential to the reel, isn't it? Unwind some very heavy cable from a drum and pull. The unwound cable need not be tangential; its resistance to bending may be too much for your strength. A perfect chain is perfectly flexible. We conclude that the tension in the chain is everywhere tangential to it.

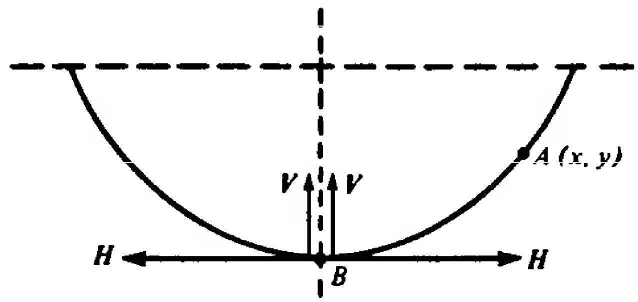

Figure 5.12

Partial corroboration of this conclusion is given by consideration of the forces acting at B, the bottom of the curve. The crux of the matter is that B is symmetrically placed with respect to the left- and right-hand portions of the curve, so that no matter what forces are exerted on it by the one portion, symmetrically identical forces are exerted on it by the other. Since H stands for *Horizontal* and V for *Vertical* let us suppose H to be the horizontal and V the vertical (upward) component of the pull exerted on B by either portion. See Fig. 5.12. The symmetry is such that the horizontal components H are equal but opposite; each annuls the effect of the other. Contrariwise, the symmetry is such that the vertical components V have a resultant $2V$. For equilibrium $2V$ must be annulled by the weight of the particle at B. What is its weight? Since the chain weighs λ per unit length, the smaller the length of the particle, the less its weight. But isn't it singularly odd to speak of the length of a particle? Arbitrarily short, it weighs arbitrarily little, so that $2V$, and consequently V, is arbitrarily small. With an ideal particle it follows that $V = 0$; the only force exerted by either portion of the chain is horizontal. But the tangent at B is horizontal. We again conclude that the tension at B is tangential. If you sever the chain at B, in what direction must you

pull to maintain BA in equilibrium? Surely your muscles give you the same answer.

Progress has been made: we are agreed that if the chain is in equilibrium, the tension is everywhere tangential to it. Thus, in particular, the equilibrium of BA is effected by a horizontal pull H at B and a pull T (say) at A tangential to the chain. What other forces act upon it? Yes, only its weight. Each tiny portion of the chain is subjected by gravity to a downward pull proportional to its length. Yet we do not need consider these forces individually; their combined effect is just as if the whole weight W of the chain BA were concentrated at a certain point (somewhere in the plane of, but not necessarily on, the chain). And what is this point called? Yes, the center of gravity. So BA may be regarded as in equilibrium under the action of three forces, H, T, and W. It follows that the force W must be equal and opposite to the resultant of the other two, and that the lines of action of these forces must be concurrent. Let them meet at C. (C is for *Concurrent.*) See Fig. 5.13.

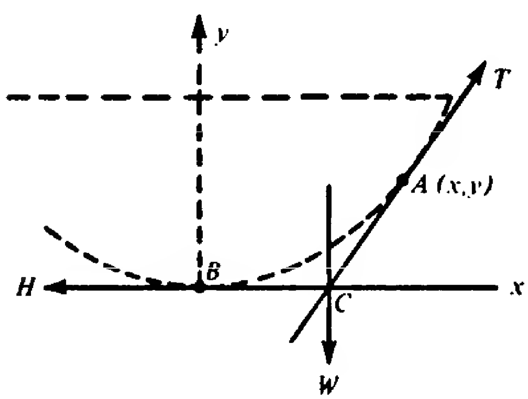

Figure 5.13

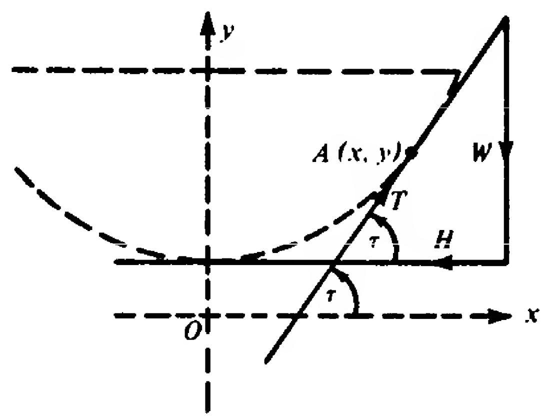

Figure 5.14

Think about this situation. With a little thought it becomes clear that we are now able to describe the shape of the curve. Our description is: no matter where the point $A(x, y)$ may be, the equation of BA is such that the line through the center of gravity of BA, parallel to the y-axis, passes through C, the point of intersection of the tangent at A and the line through B perpendicular to the y-axis. Can we obtain a differential equation from this? Not very inviting, is it? Well, perhaps there is a more amenable description. Let's try again.

Instead of starting from the fact that the lines of action of H, T, and W are concurrent, let us begin with the fact that these three forces are in equilibrium. It follows that lines representing them in magnitude as well as direction will form a closed vector triangle. Consider Fig. 5.14. As

in our first example, taking τ to be the angle made by the tangent line with the x-axis, we have immediately that

$$\tan \tau = \frac{W}{H}.$$

And again using

$$\frac{dy}{dx} = \tan \tau$$

we obtain

$$\frac{dy}{dx} = \frac{W}{H}. \tag{12}$$

Is (12) a differential equation? What do we know about H and W? H is the horizontal pull at B, see Fig. 5.12, and so H is independent of the location of $A(x, y)$; H is constant. W is the weight of the arc of the chain between point B and $A(x, y)$; if the length of this arc is s, then

$$W = \lambda s. \tag{13}$$

Our next task is to relate s to the coordinates of A. The required formula is a textbook commonplace, but perhaps you have forgotten it. Just in case, I shall derive it for you. My method will be the "nine-to-one" method. No, no, I do not mean it will take me four hours; I mean that I shall use nine parts intuition to one part logic.

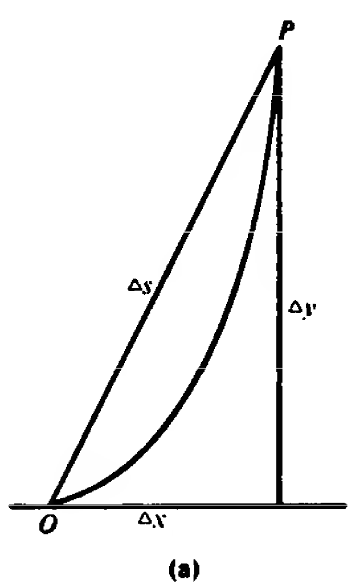

Figure 5.15(a)

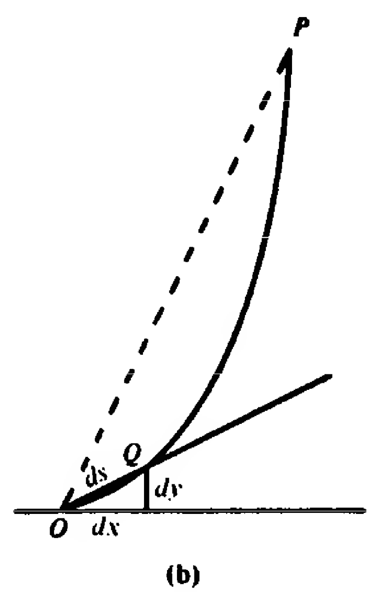

Figure 5.15(b)

Consider Figs. 5.15(a) and (b). As P moves along the curve towards O the slope of the secant OP more closely approximates that of the tangent line OQ. (We consider Q to come "infinitely close" to O.)

$$\lim_{x \to 0} \frac{\triangle y}{\triangle x} = D_x(y).$$

It is tempting to us as to Newton to describe $D_x(y)$ as the *ultimate ratio* —thereby covering up a multitude of logical sins—and to join Leibniz in writing it as dy/dx (viz., dy divided by dx). The notation

$$\lim_{x \to 0} \frac{\triangle y}{\triangle x} = \frac{dy}{dx}$$

aids and abets intuition. Both notation and diagrams tempt us to assert that when the secant reaches its limiting position, P coincides with Q, $\triangle x$ becomes dx, $\triangle y$ becomes dy, and the straight line element $\triangle s$ becomes ds, the infinitesimal bit of the curve coincident with its tangent line. Let us not resist temptation.

Bringing Pythagoras to our assistance, we find that

$$(ds)^2 = (dx)^2 + (dy)^2$$

or

$$\frac{ds}{dx} = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\,. \tag{14}$$

Yet (12) and (13) yield

$$\frac{dy}{dx} = \frac{\lambda}{H}s.$$

We differentiate this, use (14), and obtain

$$\frac{d^2y}{dx^2} = \frac{\lambda}{H}\sqrt{1 + \left(\frac{dy}{dx}\right)^2}\,. \tag{15}$$

This equation contains dy/dx and d^2y/dx^2, but no derivative of higher order than the second, so that it is a second-order differential equation and is characterized (strictly speaking, after transposition) by the general

pattern

$$F\left(x, y, \frac{dy}{dx}, \frac{d^2y}{dx^2}\right) = 0.$$

And although choice of the x-axis was left open, we nevertheless have an initial condition: at B, no matter what its ordinate, the curve is horizontal. When $x = 0$, $dy/dx = 0$.

(15) is the consequence of a physical conjecture. It ought to make sense. Does it? Let us apply the dimensions test. In an earlier application we found that dy/dx is dimensionless; schematically

$$\frac{dy}{dx} = 1,$$

so that

$$\frac{d^2y}{dx^2} = \frac{d(dy/dx)}{dx} = \frac{d(1)}{dx} = \frac{1}{L}.$$

dy/dx being dimensionless (a pure number), a bit of dy/dx, i.e., $d(dy/dx)$, is also dimensionless (a pure number). Since dy/dx is a pure number, so is $(dy/dx)^2$ and consequently also the factor $\sqrt{1 + (dy/dx)^2}$ on the right-hand side of (15). The other factor is, schematically,

$$\frac{\lambda}{H} = \frac{\text{linear density}}{\text{tension}} = \frac{\text{force/length}}{\text{force}} = \frac{1}{L}.$$

So, the right-hand side also has dimensions $1/L$. Our equation has passed its test.

Finally, for brevity, it is convenient to set

$$\frac{\lambda}{H} = \frac{1}{a},$$

where a is a length, so that (15) becomes

$$\frac{d^2y}{dx^2} = \frac{1}{a}\sqrt{1 + \left(\frac{dy}{dx}\right)^2}. \tag{16}$$

This completes the second, transitional phase.

The third, mathematical phase is to solve (16). This is a very special type of second-order equation, for it contains neither x nor y, just the first

derivative and the *derivative of the first derivative*. This very special feature enables us to do what can be done in a few more general cases: to reduce it to a first-order differential equation. You can always try; you will seldom succeed.

Let $p = dy/dx$; then

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d}{dx}(p) = \frac{dp}{dx}$$

and (16) is reduced to the first-order equation

$$\frac{dp}{dx} = \frac{1}{a}\sqrt{1+p^2}\,.$$

What are we to do next? Separate the variables, of course.

$$\int \frac{dp}{\sqrt{1+p^2}} = \frac{1}{a}\int dx. \tag{17}$$

And what do we differentiate with respect to p to get, $1/(1+p^2)^{1/2}$? Yes, an answer you should know by heart, $\log_e(p + (1+p^2)^{1/2})$. We had better have a check-up for those with weak hearts:

$$\begin{aligned}
\frac{d}{dp}\left\{\log_e\left(p+\sqrt{1+p^2}\right)\right\} &= \frac{1}{p+\sqrt{1+p^2}}\,\frac{d}{dp}\left\{p+(1+p^2)^{1/2}\right\}\\
&= \frac{1}{p+\sqrt{1+p^2}}\left\{1+\frac{1}{2}(1+p^2)^{-1/2}\frac{d}{dp}(1+p^2)\right\}\\
&= \frac{1}{p+\sqrt{1+p^2}}\left\{1+\frac{1}{2}(1+p^2)^{-1/2}2p\right\}\\
&= \frac{1}{p+\sqrt{1+p^2}}\left\{1+\frac{p}{\sqrt{1+p^2}}\right\}\\
&= \frac{1}{p+\sqrt{1+p^2}}\;\frac{p+\sqrt{1+p^2}}{\sqrt{1+p^2}}\\
&= \frac{1}{\sqrt{1+p^2}}\,.
\end{aligned}$$

Consequently, integration of (17) gives

$$\log_e\left(p + \sqrt{1 + p^2}\right) = \frac{x}{a} + C.$$

Making use of the initial condition

$$p = \frac{dy}{dx} = 0 \quad \text{when} \quad x = 0,$$

we get

$$\log_e 1 = C, \quad \text{so} \quad C = 0,$$

and we find

$$\log_e\left(\frac{dy}{dx} + \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\right) = \frac{x}{a}. \tag{18}$$

Recalling that $\log_{10}2 = 0.30103$, $2 = 10^{0.30103}$ we analogously transform (18) and obtain

$$\frac{dy}{dx} + \sqrt{1 + \left(\frac{dy}{dx}\right)^2} = e^{x/a}.$$

The next step is to get rid of the radical

$$\sqrt{1 + \left(\frac{dy}{dx}\right)^2} = e^{x/a} - \frac{dy}{dx}$$

by squaring:

$$1 + \left(\frac{dy}{dx}\right)^2 = e^{2x/a} - 2e^{x/a}\frac{dy}{dx} + \left(\frac{dy}{dx}\right)^2,$$

so that

$$2e^{x/a}\frac{dy}{dx} = e^{2x/a} - 1,$$

giving finally

$$\frac{dy}{dx} = \frac{e^{2x/a} - 1}{2e^{x/a}} = \frac{1}{2}\left(e^{x/a} - e^{-x/a}\right),$$

a first-order differential equation. Solve this and we have effected the solution of a second-order differential equation by solving successively two first-order equations.

Separating the variables, we arrive at

$$\int dy = \frac{1}{2}\int e^{x/a}dx + \frac{1}{2}\int(-e^{-x/a})dx.$$

Since the derivative of $ae^{x/a}$ is $e^{x/a}$, and the derivative of $ae^{-x/a}$ is $-e^{-x/a}$,

$$y = \frac{1}{2}ae^{x/a} + \frac{1}{2}ae^{-x/a} + C'.$$

When $x = 0$,

$$y = \frac{1}{2}a(e^0 + e^{-0}) + C' = \frac{1}{2}a(1 + 1) + C' = a + C'.$$

The simplest equation available to us is given by $C' = 0$; whereupon the equation of the catenary, the shape of a uniform chain hanging freely, is

$$y = a \cdot \frac{(e^{x/a} + e^{-x/a})}{2}. \tag{19}$$

We obtained this equation by taking y to be equal to a when $x = 0$. If now, in Fig. 5.11, we put OB equal to a, my reason for not putting the x-axis at the bottom of the curve is made apparent (as promised when Fig. 5.11 was introduced).

There remains one point to complete the solution: we have considered the shape of a uniform chain hanging freely from two supports A, A' in the same horizontal plane. What happens in an asymmetrical case where A, A' are at different levels and B no longer lies on an axis of symmetry? If the curve has a minimum point B, the bottom of a hollow, it still follows that the tension H at B is horizontal, so that the rest follows as before, despite asymmetry.

There is an alternative argument which holds even if the chain is so short that it has no bottom-of-a-hollow point B. Consider Fig. 5.16.

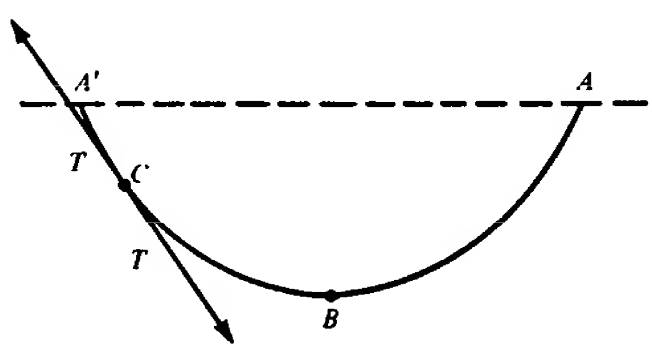

Figure 5.16

Suppose someone cuts away part AC of the chain, but holds the remaining part $A'C$ at C, supplying the tension T that AC exerted on $A'C$. The $A'C$ part is under the influence of the same forces as before and so keeps the shape it had before. Since C is an arbitrary point, (19) covers all possible cases.

5.1.4 Fall with Friction

There is some analogy between learning how to solve problems of science by mathematics and learning a foreign language: an analogy close enough to merit remark. Consider the English-speaking person who decides to learn French. At first, our would-be linguist has to do all his thinking in English; having decided what he wishes to say, he has to wrestle with the problem of translating his English into French. Later, with increased facility, he is often able to reply in French to a question asked in French without intermediary translation of the question into English and his answer to it into French. Finally, given ability and perseverance, he seldom if ever needs to fall back on his English; he thinks in French. Somewhat analogously, with scientific problems, given mathematical facility, there is no need for prior painstaking formulation in English of the physical condition; it can be expressed in mathematical notation directly. Phases 1 and 2 may go forward, as the animals went into the Ark, paired; they need not march in single file.

Our next problem is to specify the motion of a body let fall from rest in a resisting medium; for example, that of a stone when air resistance is taken into account. Let us deal conjointly with a conjecture of physical condition and its mathematical formulation; we have the facility.

We consider the fall of a particle of mass m. To provide a convenient frame of reference for our observations, we introduce an x-axis vertically downwards such that the particle is let fall from the origin when we start our stop watch. Thus in addition to the initial condition given by our coordinate system that $x = 0$ when $t = 0$, we also have a physical condition, $dx/dt = 0$. Obviously, the position of the falling particle is dependent upon the time for which it has been falling; x depends upon t, say $x = f(t)$. Our problem is to specify $f(t)$. See Fig. 5.17.

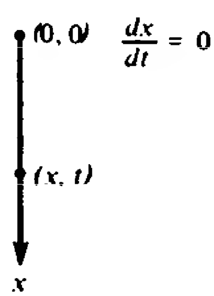

Figure 5.17

Our difficulty is to conjecture the condition upon which the motion depends. Since

$$\text{mass} \times \text{acceleration} = \text{force},$$

a stone of mass m falling under gravity, without friction, satisfies the condition

$$m \cdot \frac{d^2x}{dt^2} = mg.$$

What modification does air resistance introduce? Not only is the problem difficult for us, it is by and large an unsolved problem. Knowledge of friction has little theoretical foundation. However, we all know as a matter of crude observation that friction opposes the motion. So, if R is the frictional force opposing the motion of our particle m, we have

$$m \cdot \frac{d^2x}{dt^2} = mg - R.$$

On what does R depend? It has been found experimentally that R increases as the velocity v of the falling body increases. The simple assumption that R is directly proportional to v turns out to be too small an estimate; the assumption that R is proportional to v^2, too large. Although wrong, the latter more closely fits the facts. Thus a better assumption would appear to be that R is proportional to v^α where $1 < \alpha < 2$. In consequence, the condition for the motion is

$$m \cdot \frac{d^2x}{dt^2} = mg - Kv^\alpha,$$

where the constant of proportionality K is positive. The best empirical value of α is about 1.71. It is not theoretically conclusive.

To obtain a differential equation amenable to simple treatment we shall take α to be 1. And since $v = dx/dt$, we have

$$m \cdot \frac{d^2x}{dt^2} = mg - K\frac{dx}{dt}.$$

Dividing by m, we obtain

$$\frac{d^2x}{dt^2} = g - \frac{K}{m}\frac{dx}{dt}.$$

For brevity, we put $K/m = k$, a positive number, so that the condition for the motion is

$$\frac{d^2x}{dt^2} = g - k\frac{dx}{dt}. \tag{20}$$

The physical conceptions involved suggest the substitutions

$$\frac{dx}{dt} = v, \qquad \frac{d^2x}{dt^2} = \frac{d}{dt}\left(\frac{dx}{dt}\right) = \frac{d}{dt}(v) = \frac{dv}{dt}.$$

With these substitutions (20) becomes

$$\frac{dv}{dt} = g - kv.$$

We have a first order differential equation that cries out for separation of its variables. Separation gives

$$\int \frac{dv}{g - kv} = \int dt.$$

And since

$$\frac{d}{dv}\log_e(g - kv) = -k \cdot \frac{1}{g - kv},$$

integration yields

$$-\frac{1}{k}\log_e(g - kv) = t + C,$$

and

$$\log_e(g - kv) = -kt - kC.$$

Again recalling $\log_{10}2 = 0.30103$, etc., we have

$$g - kv = e^{-kt-kC}.$$

Transposing g and dividing by $-k$, we get

$$v = \frac{g}{k} - \frac{1}{k}e^{-kt-kC}. \tag{21}$$

Using the initial condition $v = dx/dt = 0$ when $t = 0$, (21) gives

$$0 = \frac{g}{k} - \frac{1}{k}e^{-k0}\cdot e^{-kC} = \frac{g}{k} - \frac{1}{k}e^{-kC},$$

so that

$$g = e^{-kC},$$

and

$$v = \frac{g}{k} - \frac{g}{k}e^{-kt}. \tag{22}$$

But since $dx/dt = v$, this becomes

$$\frac{dx}{dt} = \frac{g}{k} - \frac{g}{k}e^{-kt}. \tag{23}$$

Again we have reduced a second-order differential equation to two consecutive first-order equations. Unfortunately this is not always possible.

Separating the variables in (23), we get

$$\int dx = \int\left(\frac{g}{k} - \frac{g}{k}e^{-kt}\right)dt,$$

so that

$$x = \frac{g}{k}t + \frac{g}{k}\cdot\frac{e^{-kt}}{k} + C'.$$

And by the initial condition $x = 0$ when $t = 0$,

$$0 = 0 + \frac{g}{k^2}\cdot e^{-k0} + C' = \frac{g}{k^2} + C',$$

$$-\frac{g}{k^2} = C',$$

so that

$$x = \frac{g}{k}t + \frac{g}{k^2}e^{-kt} - \frac{g}{k^2}. \tag{24}$$

We have determined $f(t)$; we have a formula for free fall with friction.

Does it fit the facts? To check quantitatively is a matter for the laboratory and costs time and money, so first check qualitatively. Often much can be checked in this way with little computation. We begin with the obvious question: How is our formula related to Galileo's formula for free fall without friction,

$$x = \tfrac{1}{2}gt^2?$$

When there is no friction, $K = 0$, so that $k = 0$, and (24) should reduce to Galileo's. But we cannot substitute $k = 0$ because we cannot divide by zero; and it is difficult to see what happens when k is close to zero. We will have to postpone this check until the next section (Number 5.2.2).

But (24) is derived from (22), so that it can be checked indirectly by checking the latter. With $k > 0$, as t increases, e^{-kt} (that is $1/e^{kt}$) tends to zero; so that, by (22),

$$\text{as } t \text{ increases, } v \text{ tends to } \frac{g}{k}.$$

Thus there is a *terminal velocity* in the sense that no matter for how long the body is falling its velocity will not exceed g/k. A terminal velocity is confirmed by the experience of parachutists jumping from high altitudes —and has yet to be denied by those jumping without chutes. Chemists observe that particles sinking in a strongly viscous fluid quickly acquire a velocity that remains sensibly constant. To this small extent at least our conjectured formula is in accordance with the facts.*

SECTION 2. APPROXIMATE FORMULAE: POWER SERIES

Introduction

In most applications of mathematics to science approximate formulae play a role. Often it turns out that the full solution is too complicated or even inaccessible. When we cannot obtain the exact answer we must content ourselves with the next best thing, a good approximation. Yet the situation is not really as bad as, at first sight, it seems. Usually, provided we are energetic enough to perform the labor of calculation, we can obtain a numerical answer correct to as many places of decimals as we

*We have used the separation of variables in four cases (subsections 5.1.1–5.1.4) and we shall use it in two more (5.2.4 and 5.4.1); but, of course, it is not applicable to all cases, but only to a rather narrow subclass of differential equations of first order.

That great teacher Leonhard Euler wrote the first textbook of the calculus. An extensive part of it deals with differential equations. The very first chapter of this part is devoted to the separation of variables; see Euler's *Opera Omnia*, series 1, vol. 11, p. 257. Even in our times there are good reasons to begin an elementary introduction to differential equations with the separation of variables.

please. Nowadays with electronic computers to do our hard work for us, calculation is no problem. The essential tool with which increasingly exact approximations are obtained, no matter whether driven by electronics or by brain power, is a *power series*. No pun is intended.

What is a power series? Essentially, the expansion of a function of x, say, in terms of powers of x. And what has this to do with increasingly better approximations to the exact value? The basic idea is illustrated by a sequence of successively better values of π.

$$\begin{array}{l}
3\\
3.1\\
3.14\\
3.14\ 1\\
3.14\ 15\\
3.14\ 159\\
3.14\ 159\ 2\\
3.14\ 159\ 26\\
3.14\ 159\ 265\\
3.14\ 159\ 265\ 3\\
3.14\ 159\ 265\ 35
\end{array}$$

(We add, parenthetically, that there exists in French a delightful mnemonic for π, a poem of which the number of letters in the n^{th} word is the n^{th} figure of the decimal expansion of π. The first line of this poem runs:

Que	j'	aime	à	faire	apprendre	un	nombre	utile	aux	sages
3	.1	4	1	5	9	2	6	5	3	5 .)*

To expose the idea we write our last approximation for π with a change of emphasis.

$$\pi \approx 3 + 1\left(\frac{1}{10}\right) + 4\left(\frac{1}{10}\right)^2 + 1\left(\frac{1}{10}\right)^3 + 5\left(\frac{1}{10}\right)^4 + 9\left(\frac{1}{10}\right)^5$$

$$+ 2\left(\frac{1}{10}\right)^6 + 6\left(\frac{1}{10}\right)^7 + 5\left(\frac{1}{10}\right)^8 + 3\left(\frac{1}{10}\right)^9 + 5\left(\frac{1}{10}\right)^{10}.$$

We have an approximation for π expressed as a power series, as a series of powers of $(1/10)$. To determine the first $n + 1$ figures of the decimal

*For more digits in the decimal expansion consider the continuation

Immortel Archimede, artiste, ingénieur

Qui de ton jugement peut priser la valeur!

and for a mnemonic in English, consider Eddington's

How I need a drink, alcoholic of course, after the heavy chapters involving quantum mechanics.

expansion of π is to find the coefficients $a_0, a_1, a_2, \ldots, a_n$ such that

$$\pi = a_0 + a_1\left(\frac{1}{10}\right) + a_2\left(\frac{1}{10}\right)^2 + a_3\left(\frac{1}{10}\right)^3 + \cdots + a_n\left(\frac{1}{10}\right)^n + \cdots.$$

The more coefficients we compute, the more accurately we determine π; by finding a sufficient number of coefficients we find π with whatever accuracy we please.

Doesn't the basic idea speak for itself? Replace the powers of $(1/10)$ by powers of x and we have a function of x:

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots + a_nx^n + \cdots.$$

In this expansion, however, we must be prepared to accept for the coefficients $a_0, a_1, a_2, \ldots, a_n, \ldots$ numerical values of any kind (and not only the digits $0, 1, 2, \ldots, 9$). Our optimistic conjecture is that any function $f(x)$ can be expanded as a power series of x such that the more coefficients $a_0, a_1, a_2, \ldots, a_n$ we use, the more accurately we can determine $f(x)$ for any given numerical value of x.

It turns out that our optimism is well rewarded; almost all the useful functions can be expanded in this way. There is just one complication: the requirement that accuracy increase with more and more terms usually places a restriction on the numerical values of x for which the power series expansion will work. However, there are methods of circumventing this restriction, as we shall presently illustrate in Number 5.2.1.

We conclude this introduction by listing a few well-known power series expansions.

(a) Some expansions that hold without numerical restriction on x:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$

(b) Some expansions that require numerical restriction on x:

$$\log_e(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \qquad (-1 < x < 1)$$

$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots \qquad (-1 < x < 1)$$

$$\log_e x = \frac{x-1}{x} + \frac{1}{2}\left(\frac{x-1}{x}\right)^2 + \frac{1}{3}\left(\frac{x-1}{x}\right)^3 + \cdots \qquad \left(x > \frac{1}{2}\right)$$

5.2.1 Calculation of $\sqrt[3]{28}$

The reader will likely complain that our little list of power series omits the best-known one of all: the binomial theorem, discovered by Newton while still an undergraduate at Cambridge. It is

$$(1 + x)^a = 1 + \frac{ax}{1!} + \frac{a(a-1)x^2}{2!} + \frac{a(a-1)(a-2)x^3}{3!} + \cdots .$$

We did not list it under (a) because in certain circumstances there is a restriction on x; we did not list it under (b) because in certain circumstances there is no restriction on x; it all depends on a. If a is a positive integer, say n, the exact value of $(1 + x)^n$ for any given numerical value of x can, of course, be readily calculated. The necessity to content ourselves with an approximation does not arise. And since there is no necessity for approximation, there is, *a fortiori*, no need for a restriction on x to give arbitrarily close approximations.

If a is not a positive integer, the expansion of $(1 + x)^a$ has no last term. And since there is no last term we cannot sum them all by adding successive terms one at a time; we could never finish. When the exact value is inaccessible we must content ourselves with an approximation. It turns out that the restriction for arbitrarily close numerical approximations is that x be numerically less than 1.

How is it that we sometimes need and sometimes do not need a restriction on the numerical value of x when the expansion has no last term? Suppose that we have an expansion for π of which the first four terms are

$$3, \frac{1}{10}, \frac{4}{10}, \frac{-4}{10}.$$

These give rise to the following successive approximations

$$3, 3.1, 3.5, 3.1.$$

The fourth approximation is only as good as the second and the third is worse. In fact the third is even worse than the first. Doesn't this expansion make it obviously desirable for practical computation to place some restriction on the relative size of successive terms? Ideally, we require an expansion such that after the first few terms each term is only a fraction of its predecessor, so that later terms in the expansion can be neglected without serious error. Isn't it obvious that the SMALLER we make x in the binomial expansion of $(1 + x)^a$, the fewer terms we need take into account to get accuracy to, say, five decimal places? Determination of the largest x for which any given power series meets our requirement is difficult. The complete answer is the theory of convergence. It answers our purpose to be told if a restriction on x is necessary, and when

necessary, to be told what the restriction is. We have already given examples with restricted x in (b) above. It remains to ask: Why do some series, for example those of (a), meet our requirement without restriction on x? With such series our requirements are, so to speak, already built in; it so happens that no matter how large a numerical value is given to x, a stage will be reached in the computation after which each term is only a fraction of its predecessor. If x is small this stage is reached after a few terms; the larger x, the later this stage and the more laborious the computation for the same accuracy.

We first illustrate the utility of power series by using the binomial theorem to compute the cube root of 28. How are we to apply it? Take another look; it is stated above.

$$\sqrt[3]{28} = (28)^{1/3}$$

so that $a = \frac{1}{3}$. Also, $28 = 1 + x$, so that $x = 27$. We have

$$\sqrt[3]{28} = (1 + 27)^{1/3}.$$

But since a is not a positive integer, there is, we recall, a restriction on x: x must be numerically less than 1, which 27, alas, is not. Earlier, we remarked that there is sometimes a way of circumventing the restriction on x. As circumventing the restriction on liquor during prohibition, it requires a little ingenuity. What is an approximation to $\sqrt[3]{28}$? Yes, a little more than 3. Why a little more? A little more than 3 because $3^3 = 27$. And doesn't this suggest writing the following?

$$1 + 27 = 27\left(\frac{1}{27} + 1\right),$$

so that

$$\sqrt[3]{28} = \left\{27\left(1 + \frac{1}{27}\right)\right\}^{1/3} = 27^{1/3}\left(1 + \frac{1}{27}\right)^{1/3} = 3\left(1 + \frac{1}{27}\right)^{1/3}.$$

Now the x of $(1 + x)^{1/3}$ is $1/27$ and complies with the restriction. It is SMALL relative to unity. This is important, so we write the word in large letters. Why important? Important becasue we can anticipate soon reaching a stage where a term becomes only a SMALL fraction of its predecessor, thereby indicating that we can obtain given accuracy with relatively little labor. SMALL x gives the uplift of a hymn in praise of idleness.

Applying the binomial expansion with $a = 1/3$, $x = 1/27$, we have

$$\begin{aligned}
\sqrt[3]{28} &= 3\left(1 + \frac{1}{27}\right)^{1/3} \\
&= 3\left\{1 + \frac{\frac{1}{3}}{1!}\left(\frac{1}{27}\right) + \frac{\frac{1}{3}\left(\frac{1}{3} - 1\right)}{2!}\left(\frac{1}{27}\right)^2 \right. \\
&\qquad\qquad \left. + \frac{\frac{1}{3}\left(\frac{1}{3} - 1\right)\left(\frac{1}{3} - 2\right)}{3!}\left(\frac{1}{27}\right)^3 + \cdots \right\} \\
&= 3\left\{1 + \frac{1}{3}\left(\frac{1}{27}\right) - \frac{2}{9}\cdot\frac{1}{2!}\left(\frac{1}{27}\right)^2 + \frac{10}{27}\cdot\frac{1}{3!}\left(\frac{1}{27}\right)^3 + \cdots\right\} \\
&= 3\left\{1 + \frac{1}{3}\left(\frac{1}{27}\right) - \frac{1}{9}\left(\frac{1}{27}\right)^2 + \frac{5}{81}\left(\frac{1}{27}\right)^3 + \cdots\right\} \\
&= 3 + \left(\frac{1}{27}\right) - \frac{1}{3}\left(\frac{1}{27}\right)^2 + \frac{5}{27}\left(\frac{1}{27}\right)^3 + \cdots \\
&= 3
\end{aligned}$$

$+0.037\ 037\ \ldots$	$(\approx +\frac{1}{27})$
$-0.000\ 457\ \ldots$	$(\approx -\frac{1}{3}(\frac{1}{27})^2)$
+ less than $0.000\ 010$	$(\approx +\frac{5}{27}(\frac{1}{27})^3)$
$-0.00\ 000\ \ldots$	($\approx$ 4th and following terms).

Isn't it jolly? Right from the beginning every term is only a small fraction of its predecessor, and the farther we expand the smaller the small fraction becomes. Where we stop depends upon the accuracy we require. The third and succeeding terms do not affect the first 2 places of decimals, so that merely the first two terms gives us $\sqrt[3]{28}$ correct to 2 places of decimals, namely 3.03. The fourth and succeeding terms do not affect the first 3 places, so that the first three terms give us the cube root correct to 3 places, namely 3.036. Using the fourth term, we obtain the cube root 3.0365, correct to 4 places.

The method is of course applicable to other cube roots. For example,

$$\sqrt[3]{45} = (27 + 18)^{1/3} = \left\{27\left(1 + \frac{2}{3}\right)\right\}^{1/3} = 27^{1/3}\left(1 + \frac{2}{3}\right)^{1/3}$$
$$= 3\left(1 + \frac{2}{3}\right)^{1/3}.$$

Here x is not so small, $2/3$ instead $1/27$ (18 times as big), so we can expect a little more work for the same accuracy. And we hardly need add that the method is not confined to the extraction of cube roots. One more example will suffice to show that the same little ingenuity still works. We consider $\sqrt[5]{239}$. What is the best integral approximation?

$3^5 = 243$, too big. $\qquad$ $2^5 = 32$, much too small.

(Since 3^5 is too big, 4^5 is bigger still; and 1^5 is too small.) Without doing any more arithmetic it is clear that 3 is the best integral approximation. So, we proceed thus:

$$239 = 243 - 4 = 243\left(1 - \frac{4}{243}\right) = 3^5\left(1 - \frac{4}{243}\right),$$

so that

$$\sqrt[5]{239} = \left\{3^5\left(1 - \frac{4}{243}\right)\right\}^{1/5} = (3^5)^{1/5}\left(1 - \frac{4}{243}\right)^{1/5} = 3\left(1 - \frac{4}{243}\right)^{1/5},$$

and the stage is set for a binomial performance.

Note that for cube roots, if x is very small,

$$(1 + x)^{1/3} \approx 1 + \frac{1}{3}x.$$

For example,

$$\sqrt[3]{1001} = 10\left(1 + \frac{1}{1000}\right)^{1/3} \approx 10\left(1 + \frac{1}{3}\cdot\frac{1}{1000}\right) \approx 10.003.$$

More generally, for very small x

$$(1 + x)^a \approx 1 + ax.$$

Oh yes, binomial expansions have great practical importance.

5.2.2 Fall with Friction Again

That a man does not speculate upon the outcome of his investigation is a sure sign that he has no genuine interest in it. If genuinely interested, he cannot prevent himself from forming some idea of the answer to his problem at the outset, or, subsequently, from checking his answer when he gets one.

Earlier, we had an idea—a good idea, even though obvious—for the checking of our free fall with friction formulae (22), (24). When the frictional force becomes zero, the formulae for the velocity v and the displacement x should reduce to Galileo's free fall formulae

$$v = gt$$

$$x = \tfrac{1}{2}\, gt^2.$$

Put alternatively, when air resistance is taken into account, Galileo's formulae need correction. With air resistance a falling body is retarded, it does not fall so fast. We conjecture a corrective factor that diminishes v —and of course the diminution will depend on t. We conjecture

$$v = gt - \text{correction}$$

where

$$\text{correction} = \text{a positive function of } t.$$

And if a body does not fall so fast, it does not fall so far. Likewise, for the displacement x, we conjecture

$$x = \tfrac{1}{2}\, gt^2 - \text{correction},$$

where this correction is also a positive function of t.

Yes, we had a good idea; the defect was our inability to apply it. Times have changed; power series give us the ability. Now, we can handle (22), (24). Let us do so.

We begin with (22). Substituting $-kt$ for x in the expansion for e^x given in (a), we obtain

$$(25) \qquad e^{-kt} = 1 - \frac{kt}{1!} + \frac{k^2t^2}{2!} - \frac{k^3t^3}{3!} + \frac{k^4t^4}{4!} - \dots,$$

and since there is no restriction on the numerical value of x neither is there a restriction on kt. Thus, no matter what the value assigned to kt,

by (22)

$$v = \frac{g}{k} - \frac{g}{k}\left(1 - \frac{kt}{1} + \frac{k^2t^2}{2!} - \frac{k^3t^3}{3!} + \frac{k^4t^4}{4!} - \dots\right).$$

Multiplying out the first two terms of the bracket, we have

$$v = \frac{g}{k} - \frac{g}{k} + gt - \frac{g}{k}\left(\frac{k^2t^2}{2!} - \frac{k^3t^3}{3!} + \frac{k^4t^4}{4!} - \dots\right),$$

and taking kt as a factor from this bracket,

$$v = gt - \frac{g}{k}\cdot kt\left(\frac{kt}{2!} - \frac{k^2t^2}{3!} + \frac{k^3t^3}{4!} - \dots\right),$$

so that

$$(26) \qquad v = gt - gt\left\{\frac{1}{2!}(kt) - \frac{1}{3!}(kt)^2 + \frac{1}{4!}(kt)^3 - \dots\right\}.$$

We have obtained a power series formula for v.

(26) is interesting as well as complicated. It merits careful consideration. First note that when $k = 0$ every term in the curly bracket is zero, so that when there is no friction this formula reduces to Galileo's formula

$$v = gt.$$

We confirm our first anticipation. Second, compare the power series in the curly bracket with the expansion of $(1 + 1/27)^{1/3}$ given above. Isn't there a striking similarity? Here we have powers of kt instead of powers of $1/27$. True (26) holds no matter what the magnitude of kt—but wouldn't it be nice if kt were SMALL? Now as a matter of experimental fact k is very small indeed, so that kt is small if t is not large. And since each term of the curly bracket is only a fraction of kt times its predecessor, with kt much less than 1, each term is only a fraction of a small fraction of its predecessor. Doesn't this consideration invite comparison with the fact that each term of $(1 + 1/27)^{1/3}$ is numerically less than $1/27$ of its predecessor? What must we conclude? That the higher powers of kt can be neglected without much loss of precision; that

$$v = gt - gt\left\{\tfrac{1}{2}(kt) + \text{practically nothing}\right\},$$

i.e.,

$$v = gt - \tfrac{1}{2}\,gkt^2$$

is a good approximation. The smaller kt, the better the approximation, of course. And we confirm our conjecture

$$v = gt - \text{correction}$$

where

$$\text{correction} = \text{a positive function of } t.$$

It is worth remarking that even if kt, though small, were insufficiently small for the accuracy we require to neglect the second and some higher powers of kt, (26) would still confirm this conjecture. The reason is not far to seek. Consider the terms in the curly bracket to be paired thus:

$$(27) \qquad \left\{ \left[\frac{1}{2!}(kt) - \frac{1}{3!}(kt)^2 \right] + \left[\frac{1}{4!}(kt)^3 - \frac{1}{5!}(kt)^4 \right] + \ldots \right\}.$$

Provided $0 < kt < 3$, so that $1/3$ of, and smaller fractions of, kt are less than 1,

$$\left[\frac{1}{2!}(kt) - \frac{1}{3!}(kt)^2 \right] = \frac{1}{2}kt\left(1 - \frac{1}{3}kt\right) = \text{positive quantity}$$

$$\left[\frac{1}{4!}(kt)^3 - \frac{1}{5!}(kt)^4 \right] = \frac{1}{4!}kt\left(1 - \frac{1}{5}kt\right) = \text{positive quantity}$$

and similarly for succeeding pairs. Thus the curly bracket in (26) is still a positive quantity and consequently the correction is still a positive function of t.

Next, in essentially the same way we deal with (24). Substituting (25), we get

$$x = \frac{g}{k}t - \frac{g}{k^2} + \frac{g}{k^2}\left(1 - \frac{kt}{1} + \frac{k^2t^2}{2!} - \frac{k^3t^3}{3!} + \frac{k^4t^4}{4!} - \frac{k^5t^5}{5!} + \ldots\right)$$

$$= \frac{g}{k}t - \frac{g}{k^2} + \frac{g}{k^2} - \frac{g}{k}t + \frac{1}{2}gt^2 + \frac{g}{k^2}\left(-\frac{k^3t^3}{3!} + \frac{k^4t^4}{4!} - \frac{k^5t^5}{5!} + \ldots\right),$$

so that, simplifying, we have

$$x = \frac{1}{2}gt^2 + \frac{g}{k^2}\left(-\frac{k^3t^3}{3!} + \frac{k^4t^4}{4!} - \frac{k^5t^5}{5!} + \ldots\right).$$

Taking $-k^2t^2$ as a factor from this bracket, we get

$$x = \frac{1}{2}gt^2 - \frac{g}{k^2}\cdot k^2t^2\left(+\frac{kt}{3!} - \frac{k^2t^2}{4!} + \frac{k^3t^3}{5!} - \dots\right),$$

so that

$$(28) \quad x = \frac{1}{2}gt^2 - gt^2\left\{\frac{1}{3!}(kt) - \frac{1}{4!}(kt)^2 + \frac{1}{5!}(kt)^3 - \dots\right\}.$$

We have obtained a power series also for x.

Compare (28) with (26). How much alike can two peas from the same pod be? *Mutatis mutandis*, we draw the same conclusions. When $k = 0$ every term in the curly bracket of (28) is also zero, so that when there is no friction the formula reduces to Galileo's formula

$$x = \tfrac{1}{2}gt^2,$$

as we anticipated. Neglecting the second and higher powers of kt, we have

$$x = \frac{1}{2}gt^2 - gt^2\left\{\frac{1}{3!}(kt) - \text{practically nothing}\right\},$$

i.e.,

$$x = \tfrac{1}{2}gt^2 - \tfrac{1}{6}gkt^3$$

is a good approximation (when kt is small). We confirm our conjecture

$$x = \tfrac{1}{2}gt^2 - \text{correction},$$

where the correction is a positive function of t. It is left as an exercise for the reader to show by considering (28)'s analogue of (27), that even if kt is not sufficiently small for the accuracy we require to neglect the earliest higher powers of kt, it is still the case that the correction is positive if $0 < kt < 4$.

We cannot claim that our formulae

$$v \approx gt - \tfrac{1}{2}gkt^2$$

$$x \approx \tfrac{1}{2}gt^2 - \tfrac{1}{6}gkt^3$$

are of great practical importance. Remember that they are based on the inaccurate physical assumption that frictional resistance is directly pro-

portional to velocity, not to the 1.71 power of velocity. It was expedient for the purposes of illustration to sacrifice physical realism to mathematical simplicity. What is of great practical importance is the role of power series in the deduction of such simple, but good, approximations to such inherently complex equations as (22) and (24).

5.2.3 How Deep is a Well?

Newton was of the opinion that the solution of word problems is necessarily basic to anybody's and everybody's mathematical education. He wrote a high school textbook to support his contention. His viewpoint is not a modern one; his book is at odds with the spate of texts that currently appear: we do well to remember that Newton was a better mathematician than the best of our so often hasty contemporary authors. Here we can consider only one of Newton's well-worthwhile little problems. The reader who finds an appetite for more and has the wish, most commendable, to read Newton for himself, may be dismayed to learn that he wrote in Latin. I hasten to add that there is an English translation available: *Universal Algebra*. No "educator", or for that matter, educator, should be licensed to banish word problems from the curriculum until he has read Newton—in Latin.

The problem: to determine the depth of a well. The method: to time the drop of a stone into it. The crux: when the stone has gone down we have to wait for the sound to come up before we hear the splash.

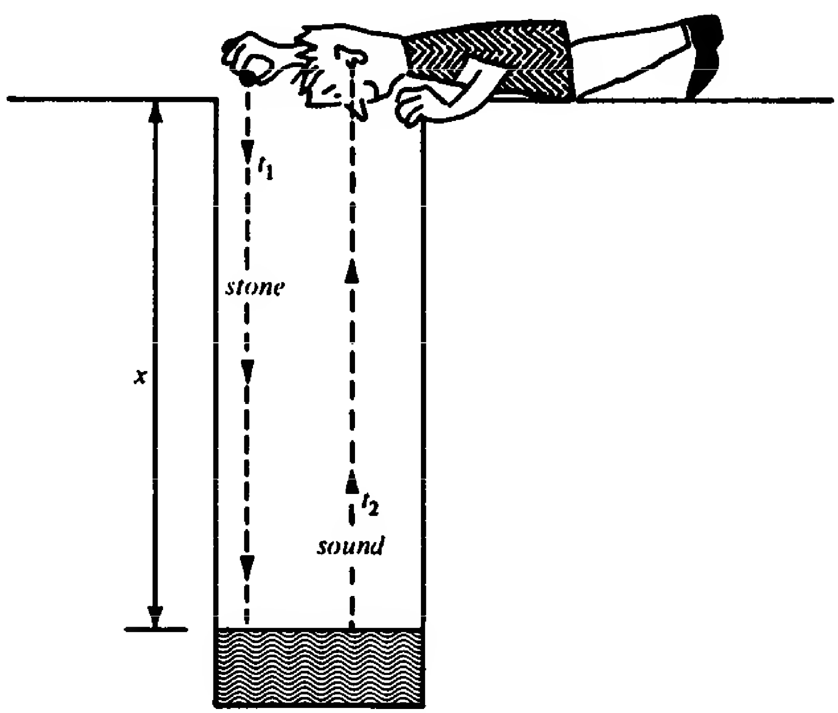

Figure 5.18

We assume a stone, a stop watch, a well—and a prone posture. See Fig. 5.18. Let t_1 be the time taken by the stone to fall a distance x, the depth of the well, and let t_2 be the time taken by the sound of the splash to rise the same distance. (As usual, we idealize; we suppose water at the bottom of our well.) The time t measured by our stop watch is the interval between dropping the stone and hearing the splash, i.e.,

$$t = t_1 + t_2.$$

With Galileo, we neglect retardation due to air resistance, so that

$$x = \tfrac{1}{2}\, g t_1^2.$$

We suppose, as is sensibly correct, the sound of the splash to be transmitted rectilinearly with uniform velocity, say c. Here c stands for *constant* and for *celeration*; sound has neither ac-celeration nor de-celeration. Consequently,

$$t_2 = \frac{x}{c}.$$

Given t, g, c our problem is to find x. As Newton is careful to explain, we have 3 unknowns, t_1, t_2, x, and 3 equations. We have as many equations as unknowns; we can hope to determine x. But we are not really interested in t_1 or t_2. So? Eliminate them. From our second equation we obtain

$$t_1 = \sqrt{\frac{2x}{g}}\,.$$

Substituting for t_1 and t_2 in our first equation, we get

$$t = \sqrt{\frac{2x}{g}} + \frac{x}{c}\,.$$

It remains to solve for x in terms of the time t measured by the stopwatch.

How are we to set about solving this equation? Take a good long look. The main point is that it contains $\sqrt{x}$ and x, i.e., $(\sqrt{x}\,)^2$. We have a quadratic in $\sqrt{x}$ in a slightly disguised form. We remove the mask:

$$\frac{1}{c}(\sqrt{x}\,)^2 + \sqrt{\frac{2}{g}}\,(\sqrt{x}\,) = t.$$

Find $\sqrt{x}$ and we can find x. To solve this quadratic we have a choice: parrots'-food formula or common-sense completion of the square. The reader may have forgotton the formula, but surely he cannot lack common sense. We complete the square. Before taking half the coefficient of $\sqrt{x}$ it is convenient to introduce a factor 2 into this coefficient by multiplying by $\sqrt{2}/\sqrt{2}$;

$$\frac{1}{c}(\sqrt{x})^2 + 2\frac{1}{\sqrt{2g}}(\sqrt{x}) = t.$$

We make the coefficient of $(\sqrt{x})^2$ unity:

$$(\sqrt{x})^2 + 2\frac{c}{\sqrt{2g}}(\sqrt{x}) = ct.$$

Half the coefficient of $\sqrt{x}$ is $c/\sqrt{2g}$. We square and add to both sides:

$$(\sqrt{x})^2 + 2\frac{c}{\sqrt{2g}}(\sqrt{x}) + \frac{c^2}{2g} = \frac{c^2}{2g} + ct,$$

i.e.,

$$\left(\sqrt{x} + \frac{c}{\sqrt{2g}}\right)^2 = \frac{c^2}{2g} + ct,$$

so that

$$\sqrt{x} + \frac{c}{\sqrt{2g}} = \pm\sqrt{\frac{c^2}{2g} + ct},$$

and

$$\sqrt{x} = -\frac{c}{\sqrt{2g}} \pm\sqrt{\frac{c^2}{2g} + ct}.$$

This equation is embarrassing. Our well has only one depth, x has only one value, yet our equation gives two. We have a choice of the plus sign or the minus sign; it is a responsible choice. No, no, don't mutter under your breath; acquire the right mental habit: vary the data. We already

know the answer in a special case. If $t = 0$, the depth of the well is of course 0. Yet when the negative sign is taken, $t = 0$ results in the absurd conclusion that

$$\sqrt{x} = \frac{-2c}{\sqrt{2g}} \neq 0.$$

We take the plus. (The fact that $t = 0$ gives $x = 0$ when the positive square root is used is some check on our algebra.)

$$\sqrt{x} = -\frac{c}{\sqrt{2g}} + \sqrt{\frac{c^2}{2g} + ct}\ .$$

We have found $\sqrt{x}$.

To find x we square.

$$x = \frac{c^2}{2g} - \frac{2c}{\sqrt{2g}}\sqrt{\frac{c^2}{2g} + ct} + \frac{c^2}{2g} + ct.$$

After slight simplification

$$x = \frac{c^2}{g} + ct - \frac{2c}{\sqrt{2g}}\sqrt{\frac{c^2}{2g} + ct}\ . \tag{29}$$

This is a nasty, cumbersome formula. Can't we simplify further? The last term contains $2g$ twice, each occurrence under a root sign. Let us utilize our observation:

$$\frac{c^2}{2g} + ct = \frac{1}{2g}(c^2 + 2gct),$$

so that

$$\sqrt{\frac{c^2}{2g} + ct} = \frac{1}{\sqrt{2g}} \cdot \sqrt{c^2 + 2gct}\ ,$$

and

$$\begin{aligned} \frac{2c}{\sqrt{2g}} \cdot \sqrt{\frac{c^2}{2g} + ct} &= \frac{2c}{\sqrt{2g}} \cdot \frac{1}{\sqrt{2g}}\sqrt{c^2 + 2gct} \\ &= \frac{c}{g}\sqrt{c^2 + 2gct}\ . \end{aligned} \tag{30}$$

A little better perhaps. Wouldn't it be worthwhile to buy simplification? The price is only a small loss of precision. Isn't a good approximation formula worth its cost?

Before we can expand the radical of (30) we must have it in the form

$$\sqrt{1+y}\,, \qquad \text{where} \qquad -1 < y < 1.$$

Are we able to meet this requirement? Which term of $\sqrt{c^2+2gct}$ is to become unity? c is approximately 1100 ft/sec and g approximately 32 ft/sec^2, so that $c/2g$ is approximately 1100/64. Now if

$$t < \tfrac{1100}{64} = 17\tfrac{3}{16}\,,$$

then

$$64t < 1100; \qquad \text{i.e.,} \quad 2gt < c,$$

and

$$2gct < c^2, \qquad \text{or} \qquad \frac{2gct}{c^2} < 1.$$

But it doesn't take 17 seconds between dropping a stone and hearing the splash with ordinary common or garden wells; they are not that deep by a long chalk. If a stone takes 14 seconds to drop, by Galileo's formula it falls $\frac{1}{2}\cdot 32\cdot 14^2$, i.e., 3136 feet, so that the sound of the splash takes less than 3 seconds to come up. Our interest is in water wells, not oil wells. It is satisfactory to us to take

$$\frac{2gct}{c^2} < 1,$$

i.e., to take

$$\frac{2gt}{c} < 1.$$

It remains to manipulate the radical of (30) into the form

$$\sqrt{1+\frac{2gt}{c}}\,.$$

$$\sqrt{c^2+2gct} = \sqrt{c^2\left(1+\frac{2gct}{c^2}\right)} = \sqrt{c^2}\sqrt{1+\frac{2gt}{c}}$$

$$= c\cdot\sqrt{1+\frac{2gt}{c}}\,,$$

so that

$$\frac{c}{g}\sqrt{c^2 + 2gct} = \frac{c^2}{g}\sqrt{1 + \frac{2gt}{c}} .$$

Hence, by (30)

$$\frac{2c}{\sqrt{2g}}\sqrt{\frac{c^2}{2g} + ct} = \frac{c^2}{g}\sqrt{1 + \frac{2gt}{c}}$$

and (29) becomes

$$x = \frac{c^2}{g} + ct - \frac{c^2}{g}\sqrt{1 + \frac{2gt}{c}} . \tag{31}$$

Before expanding into a power series we are prudent to assure ourselves that we have the correct formula to expand. Of course we cannot gain absolute assurance, yet we can make a check. It is a physical problem. Does (31) have the right dimensions? We can with propriety take centimeter and second to be our units; mass is not involved. Schematically,

$$x = \text{cm}, \qquad t = \text{sec},$$

the velocity $c = \text{cm} \cdot \text{sec}^{-1}$, the acceleration $g = \text{cm} \cdot \text{sec}^{-2}$.

So schematically (31) becomes

$$\begin{aligned} \text{cm} = {} & \frac{\text{cm}^2 \cdot \text{sec}^{-2}}{\text{cm} \cdot \text{sec}^{-2}} + (\text{cm} \cdot \text{sec}^{-1})\text{sec} \\ & - \frac{\text{cm}^2 \cdot \text{sec}^{-2}}{\text{cm} \cdot \text{sec}^{-2}}\sqrt{1 + \frac{(\text{cm} \cdot \text{sec}^{-2})\text{sec}}{\text{cm} \cdot \text{sec}^{-1}}} \\ = {} & \text{cm} + \text{cm} - \text{cm}\sqrt{1+1} \\ = {} & \text{cm} + \text{cm} - \text{cm} \\ = {} & \text{cm}. \end{aligned}$$

Remember that $\sqrt{2}$, a pure number, is of zero dimensions. Dimensionally, our equation is correct. We proceed with some confidence.

It is a good mental habit to anticipate the outcome of a procedure. What result do we expect power series expansion to give? If the splash were heard when it occurred, x would be given by Galileo's formula

$$x = \tfrac{1}{2} gt^2.$$

We anticipate

$$x = \tfrac{1}{2}\, gt^2 + \text{correction.}$$

Fuller appreciation of the physical circumstances enables us to be more precise. The time t is shared between stone and sound. Since the stone doesn't fall for so long as t it doesn't fall so far as $gt^2/2$. The correction must be negative. On these occasions when we can see as well as hear the splash we know sight and sound to be almost instantaneous. The stone takes the lion's share of t. The negative correction will be small.

Knowing what to expect, we proceed. With $a = \frac{1}{2}$, the binomial theorem gives

$$(1 + x)^{1/2} = 1 + \frac{1}{2}x + \frac{\frac{1}{2}(\frac{1}{2} - 1)}{2!}x^2 + \frac{\frac{1}{2}(\frac{1}{2} - 1)(\frac{1}{2} - 2)}{3!}x^3 + \ldots$$

$$= 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \frac{5}{128}x^4 + \frac{7}{256}x^5 - \frac{21}{1024}x^6 + \ldots$$

Taking $x = 2gt/c$ to obtain the expansion of $\sqrt{1 + 2gt/c}$, we find that (31) becomes

$$x = \frac{c^2}{g} + ct - \frac{c^2}{g}\left\{1 + \frac{1}{2}\left(\frac{2gt}{c}\right) - \frac{1}{8}\left(\frac{2gt}{c}\right)^2 + \frac{1}{16}\left(\frac{2gt}{c}\right)^3 \right.$$

$$\left. - \frac{5}{128}\left(\frac{2gt}{c}\right)^4 + \frac{7}{256}\left(\frac{2gt}{c}\right)^5 - \frac{21}{1024}\left(\frac{2gt}{c}\right)^6 + \ldots\right\}.$$

Multiplying out the first three terms of the bracket, we have

$$x = \frac{c^2}{g} + ct - \frac{c^2}{g} - ct + \frac{1}{2}\, gt^2 - \frac{c^2}{g}\left\{\frac{1}{16}\left(\frac{2gt}{c}\right)^3 \right.$$

$$\left. - \frac{5}{128}\left(\frac{2gt}{c}\right)^4 + \frac{7}{256}\left(\frac{2gt}{c}\right)^5 - \frac{21}{1024}\left(\frac{2gt}{c}\right)^6 + \ldots\right\}.$$

Hence, after simplification,

$$x = \frac{1}{2} gt^2 - \frac{c^2}{g} \left\{ \frac{1}{16} \left(\frac{2gt}{c} \right)^3 - \frac{5}{128} \left(\frac{2gt}{c} \right)^4 + \frac{7}{256} \left(\frac{2gt}{c} \right)^5 - \frac{21}{1024} \left(\frac{2gt}{c} \right)^6 + \ldots \right\}.$$

Taking out $(2gt/c)^2$ as a factor from the curly bracket, we get

$$x = \frac{1}{2} gt^2 - \frac{c^2}{g} \cdot \frac{4g^2t^2}{c^2} \left\{ \frac{1}{16} \left(\frac{2gt}{c} \right) - \frac{5}{128} \left(\frac{2gt}{c} \right)^2 + \frac{7}{256} \left(\frac{2gt}{c} \right)^3 - \frac{21}{1024} \left(\frac{2gt}{c} \right)^4 + \ldots \right\},$$

so that

$$x = \frac{1}{2} gt^2 - 4gt^2 \left\{ \frac{1}{16} \left(\frac{2gt}{c} \right) - \frac{5}{128} \left(\frac{2gt}{c} \right)^2 + \frac{7}{256} \left(\frac{2gt}{c} \right)^3 - \frac{21}{1024} \left(\frac{2gt}{c} \right)^4 + \ldots \right\}.$$

This equation is rather similar to (28). To emphasize this similarity we absorb the factor 4 in the curly bracket, giving

$$x = \frac{1}{2} gt^2 - gt^2 \left\{ \frac{1}{4} \left(\frac{2gt}{c} \right) - \frac{5}{32} \left(\frac{2gt}{c} \right)^2 + \frac{7}{64} \left(\frac{2gt}{c} \right)^3 - \frac{21}{256} \left(\frac{2gt}{c} \right)^4 + \ldots \right\}. \tag{32}$$

It is instructive to compare (32) with (28). To facilitate comparison we repeat (28):

$$x = \frac{1}{2} gt^2 - gt^2 \left\{ \frac{1}{3!} (kt) - \frac{1}{4!} (kt)^2 + \frac{1}{5!} (kt)^3 - \frac{1}{6!} (kt)^4 + \ldots \right\}.$$

It's rather like meeting an old girl friend with a new hairdo. The novelty lies within the curly bracket. First, we note a different sequence of coefficients:

$$\frac{1}{4}, \frac{5}{32}, \frac{7}{64}, \frac{21}{256}, \ldots \quad \text{instead of} \quad \frac{1}{3!}, \frac{1}{4!}, \frac{1}{5!}, \frac{1}{6!}, \ldots;$$

yet the new sequence continually decreases as does the old. Second, we note powers of $2gt/c$ instead of powers of kt, yet the powers themselves are the same. And isn't one small quantity as good as another, so to speak? And what do we conclude? That it's the same old girl friend. That the numerical value of the curly bracket in (32) is positive for small $2gt/c$, confirming our expectation

$$x = \tfrac{1}{2}gt^2 - (\text{positive correction}).$$

We may with little loss of precision neglect the second and higher powers of $2gt/c$, so that

$$x = \frac{1}{2}gt^2 - gt^2\left\{\frac{1}{4}\left(\frac{2gt}{c}\right) - \text{practically nothing}\right\},$$

i.e.,

$$x = \frac{1}{2}gt^2 - \frac{1}{2}\frac{g^2t^3}{c}$$

is a good approximation (when $2gt/c$ is small).

Isn't it astonishing that two distinctly different physical problems, that of free fall with friction and that of the depth of a well, should have such similar solutions? Their similarity bears testimony to the usefulness of power series. Let us be prepared to meet mathematics that with trivial change of detail affords solution to problems from vastly different areas of physics.

5.2.4 Pendulum: Small Oscillations

Earlier, assuming that T (the full period of oscillation of a simple pendulum) is a function of l (the pendulum's length) and g (the gravitational constant), we were able to show, merely by dimensional considerations, that

$$T = c\sqrt{\frac{l}{g}},$$

where c is independent of l and of g. That we could do so much with so little is astonishing. That we could not show $c = 2\pi$ by such limited means is not astonishing. However, we now have the means: namely, that mathematical method in science which formulates the condition as a differential equation. Let us, without further ado, use it.

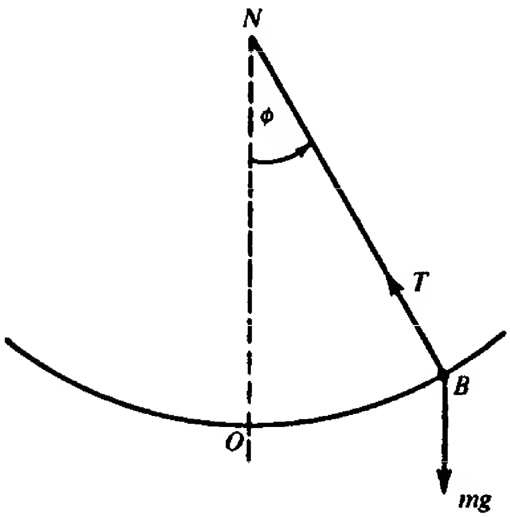

Figure 5.19(a)

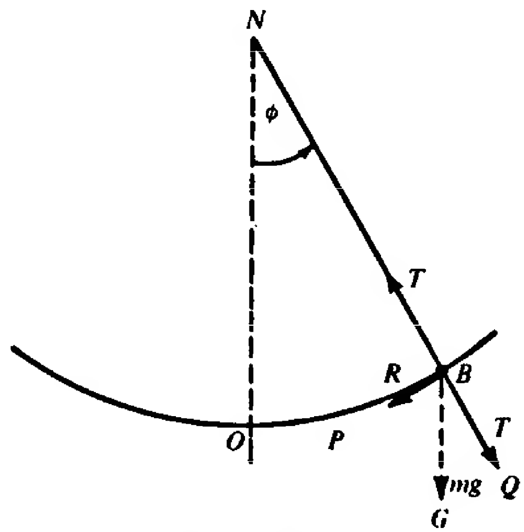

Figure 5.19(b)

What is the condition upon which the oscillation of a pendulum depends? Not so fast; we had better walk before we run. First ask: What force causes the acceleration? Consider Fig. 5.19(a). The only forces acting on the bob B, of point mass m (say), are the upward tension T in the string and mg vertically downward due to gravity. So? The accelerating force acting on B must be R, the resultant of these two. What is R? A convenient alternative to using a vector parallelogram of forces is to resolve mg into two components, the one collinear with, and the other perpendicular to, the string. See Fig. 5.19(b). Since the string is inextensible and remains taut, the component part of mg represented by $\overrightarrow{BQ}$ must be equal in magnitude and opposite to the tension T. Consequently the component of mg perpendicular to the string, represented by $\overrightarrow{BP}$ (P is for *Perpendicular*), must be the resultant R. From the obvious geometry of the figure, $\angle GBQ = \phi$, the angle made by NB with the vertical NO. And BP is perpendicular to BQ. Therefore

$$R = mg \cdot \sin \phi. \tag{33}$$

We have found the accelerating force acting on the bob.

What is a, the magnitude of the bob's acceleration? Not so fast; walk. First ask: what is its velocity? See Fig. 5.20. If in time Δt the string

turns through an increase of angle $\Delta\phi$, the bob moves along an arc length $l \cdot \Delta\phi$, so that its average velocity during this time is $l(\Delta\phi/\Delta t)$. Hence, mindful of Leibniz, we have that its instantaneous velocity is $l(d\phi/dt)$. But acceleration is rate of change of velocity, so that

$$(34) \qquad a = \frac{d}{dt}\left\{l\frac{d\phi}{dt}\right\} = l\frac{d}{dt}\left(\frac{d\phi}{dt}\right) = l\frac{d^2\phi}{dt^2}.$$

We have found the bob's acceleration.

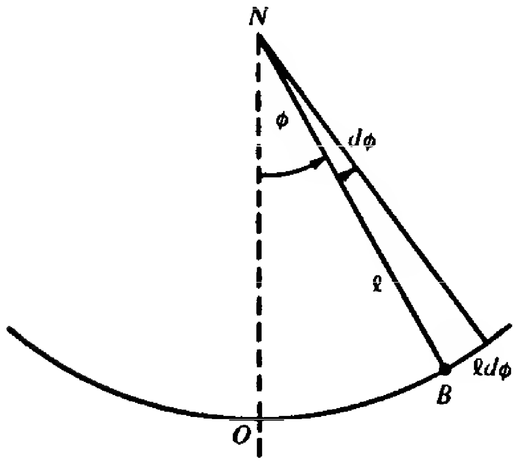

Figure 5.20

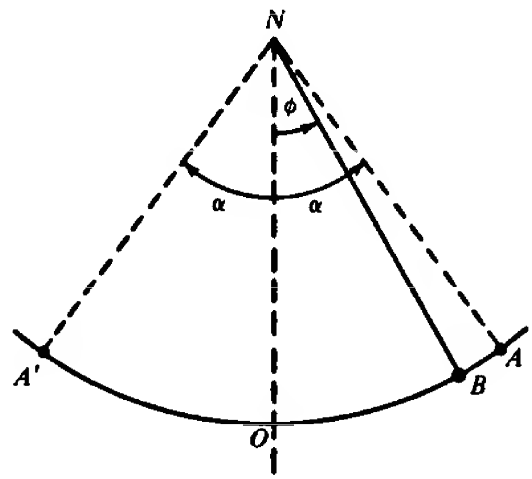

Figure 5.21

What is the relation between acceleration and accelerating force? Yes,

$$\text{mass} \times \text{acceleration} = \text{accelerating force}.$$

But be clear that it is incorrect to write

$$ma = R.$$

We have measured a along the arc in the direction of increasing ϕ and R in the opposite direction. We must measure both in the same direction. The accelerating force in the direction of increasing ϕ is $-R$; we have

$$ma = -R.$$

Substituting for a and R the expressions (34) and (33), we obtain

$$ml\frac{d^2\phi}{dt^2} = -mg\sin\phi,$$

so that

$$\frac{d^2\phi}{dt^2} = -\frac{g}{l} \cdot \sin \phi. \tag{35}$$

We have found the differential equation upon which the pendulum's oscillation is conditional.

We have yet to assign initial or boundary conditions to this differential equation. See. Fig. 5.21. We have taken the pendulum NB to be at an angle ϕ to its central position NO (where B is vertically below the nail N) at time t. When is the bob at O? It is obviously convenient to start timing the swinging pendulum from when it is in this central position. Since this is the time when we *initiate* measurements,

(i) $$\text{when } t = 0, \quad \phi = 0$$

is appropriately said to be an *initial condition*. We are agreed when our interest in the swinging pendulum begins. When does our interest end? At the end of a quarter swing, when the bob is at A; for obviously the time for the bob to swing from O to A is a quarter of the time T for a complete oscillation from O to A to O to A' to O. Thus we are led to ask: What is the value of ϕ when $t = \frac{1}{4}T$ and the bob is at A? This, the greatest value of ϕ, is said to be the *amplitude* of the oscillation. Let us call it α. So, we have,

$$\text{when} \quad t = \tfrac{1}{4}T, \quad \phi = \alpha.$$

But, is this condition genuinely informative? Would we be any the wiser if we had called the amplitude β instead of α? The giving of a name to the amplitude does not tell us anything about the amplitude itself. We are mindful of the story, possibly apocryphal, of the student who said, "Yes, yes, I understand how you determined the mass of Jupiter. What puzzles me is how you found out its name." Yes, yes, we have named the amplitude; the important thing is to determine it. This is a question of physics, not language. When the bob reaches A it is at the end of an oscillation; it is instantaneously at rest:

$$l\frac{d\phi}{dt} = 0, \qquad \text{and consequently,} \qquad \frac{d\phi}{dt} = 0.$$

Thus, we have,

(ii) $$\text{when } t = \tfrac{1}{4}T, \quad \phi = \alpha, \quad \frac{d\phi}{dt} = 0.$$

Because this condition holds when the bob reaches an end or boundary of its path, it is appropriately termed a *boundary condition*. Although A is a terminus, *terminal condition* is not accepted usage.

Now we are able to state the complete mathematical formulation of our problem:

Given (35), (i), (ii), *find* T.

Two points arise; each has a bearing on the other. The first, that solution belies brevity of formulation. As well as long, it is difficult. There are far too many mathematical difficulties for us. Exact solution involves an elliptic function, a variety distinct from the usual exponential, trigonometric, logarithmic, and algebraic expansions. Anticlimax. What are we to do? No, no, it's no use muttering. Above all, we must retain the right mental attitude. We cannot solve our problem; can we solve a simplified version? Simplicity is worth buying if we do not have to pay too great a loss of precision for it. The sensible thing to do is the next best thing; to seek a good approximation. Approximation? Approximation suggests power series expansion in powers of a small quantity.

This brings us to the second point. What sort of solution does (35) have? Since this equation contains a (second) derivative of ϕ with respect to t, we anticipate the solution to be an equation giving ϕ as a function of t (and involving the constants l, g); i.e., of the form

$$\phi = f(t, l, g). \tag{36}$$

Alternatively, consider the problem from the other end. Differentiating (36) with respect to t, we get, schematically,

$$\frac{d\phi}{dt} = f'(t, l, g),$$

and after a second differentiation with respect to t,

$$\frac{d^2\phi}{dt^2} = f''(t, l, g).$$

Either way we come to the conclusion that the solution is of the sort described by (36). Substituting the boundary condition (ii) in it, we have

$$\alpha = f(\tfrac{1}{4}T, l, g)$$

i.e., that α is given in terms of $\frac{1}{4}T$, l, and g. Hence, making T the

subject of the formula, we expect T to be given in terms of l, g, *and* α. Schematically,

$$T = F(l, g, \alpha).$$

Yet, earlier, we concluded in consequence of dimensional considerations that

$$T = c\sqrt{\frac{l}{g}},$$

where c is a constant (actually 2π). This conclusion, put schematically, is

$$T = F(l, g).$$

We concluded T to be a function of l and g *without* being a function of α. T cannot both be and not be independent of α. A dilemma confronts us.

Overcoming our despondency, we think again. The formula

$$T = c\sqrt{\frac{l}{g}}$$

was obtained by dimensional considerations *on the assumption that T is dependent upon (ONLY) l and g*. True we did not explicitly use the word *only*, true we did not explicitly state T to be independent of α; but an implicit assumption is nevertheless an assumption. We cannot quarrel with our conclusion being consistent with its premises. So the real question is: What about our premises? Is T *in fact* independent of α or not? We must resort to the final arbiter, experiment.

What is experiment's verdict? For large α it is found that T is not independent of α. When, for example, a pendulum swings with an amplitude of 60° its period is appreciably less than when it swings with an amplitude of 90°. But when α is small, say less than 10°, there is no appreciable difference in the periods of oscillation. When a pendulum does not swing so far, it does not swing so fast; decrease in arc and acceleration are compensating factors that tend to annul one another: the smaller α, the greater their annulment and the smaller the change in T; the greater α, the smaller their annulment and the greater the change in T. What are we to conclude? That although to be exact T is a function of α (as well as of l and g), if α is small its effect may be sensibly neglected.

Small α? What about ϕ? Since α is the greatest value of ϕ, when α is small, ϕ must be small. Small ϕ? The very thing for a good approximation from an expansion in powers of ϕ. And what has an expansion in powers of ϕ? Look at list (a) p. . Yes, $\sin\phi$ and $\cos\phi$. But (35) prefers $\sin\phi$. We take

$$\sin\phi = \frac{\phi}{1!} - \frac{\phi^3}{3!} + \frac{\phi^5}{5!} - \dots .$$

This expansion holds for any value of ϕ *measured in radians*. If it is not already obvious that for small values of ϕ the third and higher powers of ϕ can be neglected with little loss of accuracy, then an example will make it obvious. Take $\varphi = 10°$.

Since $180° = \pi$ radians

$$10° = \frac{\pi}{18} = 0.1745\dots \text{ radians,}$$

so that

$$\begin{aligned}\sin 10° &= 0.1745\dots - \frac{(0.1745\dots)^3}{6} + \frac{(0.1745\dots)^5}{120} - \dots \\ &= 0.1745\dots - 0.00088\dots + 0.0000013\dots - \dots .\end{aligned}$$

We conclude that for small ϕ

$$\sin\phi \approx \phi$$

with good accuracy.

Let us now use geometry to echo arithmetic. It is convenient to consider the chord and arc of a unit circle subtended by a small angle ϕ and its mirror image. See. Fig. 5.22. From the obvious geometry

$$\sin\phi = \frac{B'O}{1} = \frac{OB}{1}$$

so that

$$2\sin\phi = B'O + OB = \text{chord } B'B.$$

Since ϕ is measured in radians,

$$\text{angle} \times \text{radius} = \text{arc},$$

i.e.,

$$2\phi \times 1 = \text{arc } B'B.$$

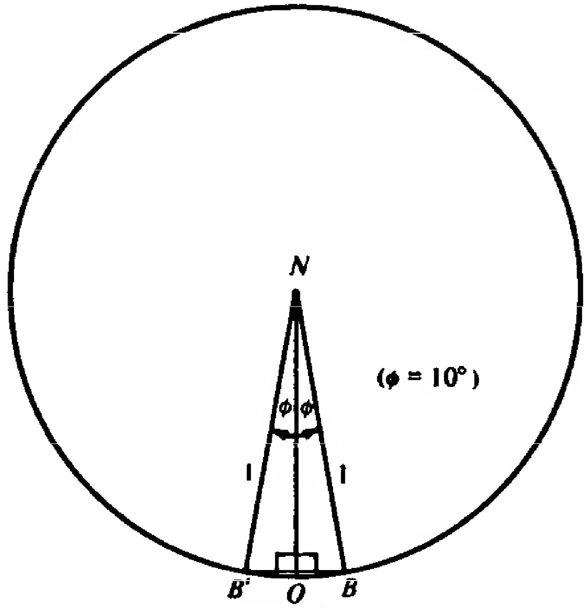

Figure 5.22

But, for small ϕ,

$$\text{chord } B'B \approx \text{arc } B'B,$$

so that

$$2 \sin \phi \approx 2\phi,$$

and

$$\sin \phi \approx \phi.$$

The smaller ϕ, the more nearly equal chord and arc; consequently the more nearly equal $\sin \phi$ and ϕ.

We conclude that for sufficiently small ϕ

$$\frac{d^2\phi}{dt^2} = -\frac{g}{l}\phi \tag{37}$$

is a good approximation to (35). In consequence we are disposed to think that the solution to (37) will be a good approximation to the solution to (35). To accept (37) as a substitute for (35) is a responsible decision; some error must be involved. How big an error? The proof of the pudding is the eating thereof; the best check is to compare the consequences of our simplified equation with the experimental facts. But before we can compare the consequences we have to deduce them.

How are we to solve our second-order differential equation? Yes, we try to reduce it to a first-order equation. And what substitution do we make? One appropriate to the physical situation. Our concern is the swing of a pendulum; rate of swing is an angular velocity, isn't it? So? We put

$$\omega = \frac{d\phi}{dt}. \tag{38}$$

Consequently

$$\frac{d\omega}{dt} = \frac{d}{dt}(\omega) = \frac{d}{dt}\left(\frac{d\phi}{dt}\right) = \frac{d^2\phi}{dt^2}$$

and (37) becomes

$$\frac{d\omega}{dt} = -\frac{g}{l}\cdot\phi. \tag{39}$$

First order? Or should we say "first disorder"? For we have three variables, ω, t, and ϕ. Two is company, three is a crowd. Who, to use a current vulgarity, is to get lost? Consider

$$\frac{d\omega}{dt} = \frac{(\ )}{dt}\cdot\frac{d\omega}{(\ )}.$$

Is $d\omega$, dt, or $d\phi$ to fill both parentheses? Try them. $d\phi$ gives

$$\frac{d\omega}{dt} = \frac{d\phi}{dt}\cdot\frac{d\omega}{d\phi}.$$

Using (38)

$$\frac{d\omega}{dt} = \omega\cdot\frac{d\omega}{d\phi}$$

so that (39) becomes

$$\omega\cdot\frac{d\omega}{d\phi} = -\frac{g}{l}\cdot\phi. \tag{40}$$

t gets lost. We have a first-order differential equation.

Next, of course we separate the variables. (40) gives

$$\int\omega\cdot d\omega = -\frac{g}{l}\int\phi\cdot d\phi.$$

Integrating, we get

$$\frac{1}{2}\omega^2 = -\frac{g}{l}\cdot\frac{1}{2}\phi^2 + c.$$

Using part of boundary condition (ii), $\phi = \alpha$, $d\phi/dt = \omega = 0$, we have

$$0 = -\frac{g}{l}\cdot\frac{1}{2}\alpha^2 + c.$$

$$c = \frac{1}{2}\,\frac{g}{l}\,\alpha^2$$

and

$$\frac{1}{2}\omega^2 = \frac{1}{2}\,\frac{g}{l}\,(\alpha^2 - \phi^2).$$

Hence, using (38) again, we obtain

$$\left(\frac{d\phi}{dt}\right)^2 = \frac{g}{l}\,(\alpha^2 - \phi^2)$$

and

$$\frac{d\phi}{dt} = \sqrt{\frac{g}{l}}\cdot\sqrt{\alpha^2 - \phi^2}\,. \tag{41}$$

We have a second first-order differential equation.

As a mere matter of routine we separate the variables. It remains to integrate

$$\int \frac{d\phi}{\sqrt{\alpha^2 - \phi^2}} = \sqrt{\frac{g}{l}}\int dt. \tag{42}$$

The left-hand side is a little awkward.

$$\alpha^2 - \phi^2 = \alpha^2\left\{1 - \left(\frac{\phi}{\alpha}\right)^2\right\},$$

so that

$$\sqrt{\alpha^2 - \phi^2} = \alpha\sqrt{1 - (\phi/\alpha)^2}$$

and

$$\int \frac{d\phi}{\sqrt{\alpha^2 - \phi^2}} = \int \frac{d\phi}{\alpha\sqrt{1 - (\phi/\alpha)^2}} = \int \frac{(1/\alpha)d\phi}{\sqrt{1 - (\phi/\alpha)^2}} .$$

But,

$$\frac{1}{\alpha}\,d\phi = d\left(\frac{\phi}{\alpha}\right)$$

so, finally

$$\int \frac{d\phi}{\sqrt{\alpha^2 - \phi^2}} = \int \frac{d(\phi/\alpha)}{\sqrt{1 - (\phi/\alpha)^2}}$$

which is of the form

$$\int \frac{dx}{\sqrt{1 - x^2}} ,$$

where

$$x = \frac{\phi}{\alpha} .$$

Since you cannot know by intuition, you must know by tuition and by heart that

$$\int \frac{dx}{\sqrt{1 - x^2}} = \text{arc sin}(x) + c.$$

Thus (42) gives

$$\text{arc sin}\left(\frac{\phi}{\alpha}\right) = \sqrt{\frac{g}{l}}\cdot t + c',$$

where the arbitrary constant of the left-hand side has been absorbed into that of the right.

Though not essential to the determination of T it is useful to have an explicit formula for ϕ. How do we get rid of *arc sin*? The relation between *arc sin of* and *sine of* is analogous to the relation between *father*

of and *son of*. Both are inverse relations. If

$$\textit{father of}\ \text{Jimmy} = \text{John}$$

then

$$\text{Jimmy} = \textit{son of}\ \text{John}.$$

Analogously, if

$$\textit{arc sin of}\ \frac{\phi}{\alpha} = A,$$

then

$$\frac{\phi}{\alpha} = \textit{sine of}\ A.$$

In consequence, our last equation gives

$$\frac{\phi}{\alpha} = \sin\left\{\sqrt{\frac{g}{l}}\cdot t + c'\right\},$$

so that

$$\phi = \alpha\cdot\sin\left\{\sqrt{\frac{g}{l}}\cdot t + c'\right\}. \tag{43}$$

It remains to determine the constant c'. The main point here is that we do not use the same condition twice. We have used the boundary condition; we now use the initial condition, (i), that $\phi = 0$ when $t = 0$. Substituting in (43), we have

$$0 = \alpha\cdot\sin\{0 + c'\}.$$

This equation has a variety of solutions; we take the simplest angle,

$$c' = 0,$$

so that

$$\phi = \alpha\sin\left\{\sqrt{\frac{g}{l}}\cdot t\right\}. \tag{44}$$

We have obtained an explicit formula for ϕ.

Finally, we are able to determine T. See Fig. 5.21 again. Because the pendulum is timed from its central position, $T/4$ is the first time at which the bob coincides with A; i.e., $T/4$ is the least value of t for which $\phi = \alpha$. Consequently, $\sqrt{g/l}\ T/4$ is the least value of $\sqrt{g/l}\ t$ for which $\phi = \alpha$. What are these values? Putting $\phi = \alpha$ in (44), we have

$$\alpha = \alpha \sin\left\{\sqrt{\frac{g}{l}}\ t\right\},$$

$$1 = \sin\left\{\sqrt{\frac{g}{l}}\ t\right\},$$

so that the values of $\sqrt{g/l}\ t$ for which $\phi = \alpha$ are

$$\frac{\pi}{2},\quad 2\pi + \frac{\pi}{2},\quad 4\pi + \frac{\pi}{2}, \ldots .$$

Thus the least is $\pi/2$. Therefore

$$\sqrt{\frac{g}{l}} \cdot \frac{1}{4} T = \frac{\pi}{2},$$

and

$$T = 2\pi\sqrt{\frac{l}{g}}\ . \tag{45}$$

We have solved the simplified version of our problem.

And now the crucial question: does the substitution of (37) for (35) result in serious error? No, it does not. With $\alpha < 10°$, there is no sensible difference between the predictions of (45) and the results of experiment. Our simplification has ample justification.

Remember that the differential equation (37),

$$\frac{d^2\phi}{dt^2} = -\frac{g}{l} \cdot \phi,$$

has the solution (44)

$$\phi = \alpha \sin\left\{\sqrt{\frac{g}{l}}\ t\right\}.$$

It is important. Important for a variety of reasons. The vibrations of tuning forks, elastic bodies, and even certain electrical phenomena are also conditional upon (37). In consequence (37) is known as the *equation of small oscillations*. You must surely meet it in physics.

SECTION 3. PHYSICAL ANALOGY

The first stage of our success in solving physical problems has been the formulation of the appropriate condition as a differential equation with an initial condition; the second, the solving of the equation subject to its initial condition. On one occasion we were unsuccessful; we could not determine the period for a pendulum with large oscillations. We may reflect that the limiting factor to our success lay in the second stage rather than the first. Even if without the Scot's proverbial thrift, the difficulty of solving differential equations is an incentive to using them parsimoniously. Happily here is a commodity of which a little may be made to go a long way. I have already made brief mention that the equation of small oscillations of a pendulum also holds for other vibrational phenomena. In investigating swinging pendulums we were, albeit unwittingly, also investigating vibrating tuning forks. Is this a straw which shows which way the wind blows? Do other differential equations have multiple uses? We have the incentive to find out.

We concern ourselves with the application of a previous result to electricity. In Number 5.1.4, Fall with Friction, we showed that the differential equation (20)

$$\frac{d^2x}{dt^2} = g - k\frac{dx}{dt},$$

with the initial condition

$$\text{when } t = 0,\ \ x = 0,\ \ \frac{dx}{dt} = 0,$$

has the consequence (22), that

$$v = \frac{g}{k} - \frac{g}{k}e^{-kt}.$$

It would seem improbable that this information could be of any interest whatsoever to the electrical engineer. Nowadays, with Telstars in regular use, intercontinental ballistic missiles ready for immediate use, and electronic computers rapidly becoming as numerous as typewriting machines,

the reader will experience no surprise when told that the study of electricity has become a most exact science. What, for goodness' sake, can an *approximate* condition for the fall of a body, dead or alive, from a hot-air balloon have to do with such an *exact* science? Life is full of surprises: our *approximate* condition for the fall of a body through a resisting medium is precisely analogous to the *exact* condition for the flow of an electric current through a resisting wire.

To be strictly correct there is a precise analogy when (20) is expressed in a fully explicit form. To gain that brevity of notation so convenient to formal manipulation we obtained (20) by dividing through by m and subsequently substituting k for K/m. To regain explicit reference to m we employ the reverse procedures in reverse order, thereby obtaining

$$m \frac{d^2x}{dt^2} = mg - K \frac{dx}{dt} .$$

And finally, since by definition

$$v = \frac{dx}{dt} ,$$

and consequently

$$\frac{dv}{dt} = \frac{d^2x}{dt^2} ,$$

we may with brevity but without any real loss of explicitness write

$$m \frac{dv}{dt} = mg - Kv. \tag{20'}$$

This is the form most convenient for making an analogy with the "fall", i.e., flow, of an electric current.

Since (20′) is explicit, the ingredients of the equivalent equation (20) are now visibly obvious; namely, in order from left to right, mass m, rate of change of velocity dv/dt, gravitational force mg, and velocity v. What are their electrical counterparts? See Fig. 5.23. To press the switch, to allow a current to start flowing is the analogue of opening the fingers, to allow a body to start falling. The fall of the body is caused by the force mg due to gravity; the flow of the current is caused by the electromotive force or tension E due to the battery. The falling body has to overcome the frictional resistance of the air; the flowing current has to overcome the electrical resistance of the wire. Air resistance is proportional to the body's velocity v; electrical resistance is proportional to the current i.

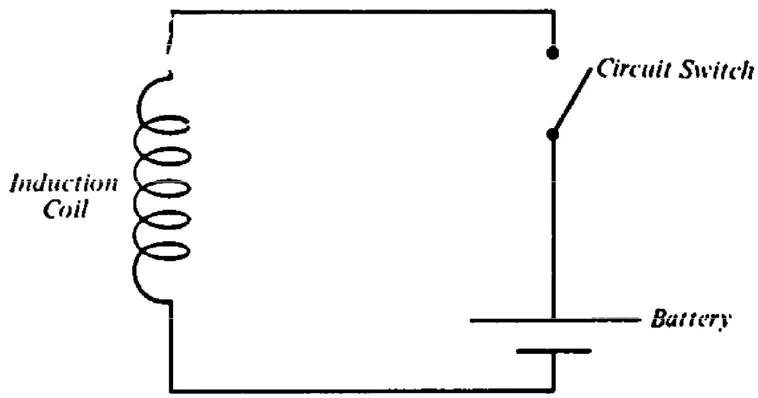

Figure 5.23

and consequently rate of change of velocity dv/dt corresponds to rate of change of current di/dt. We tabulate these analogues. "$\sim$" means *analogous to*.

Electrical	*Analogy*	*Physical*
Electromotive force	$E \sim mg$	Gravitational force
Intensity of flowing current	$i \sim v$	Velocity of falling body
Rate of change of current	$\frac{di}{dt} \sim \frac{dv}{dt}$	Rate of change of velocity
	$? \sim m$	Mass of body

We are confronted with a blank on the left-hand side. What is the analogue of mass? The electromagnetic induction L opposes change of current so that a current cannot be quite instantaneously started or stopped. And doesn't the inertia or mass m of a body tend to make it go on forever without increasing or decreasing its motion? Isn't L, so to speak, an electromagnetic inertia. We complete our list.

	Analogy	
Self-induction	$L \sim m$	Inert mass

Having found what are more or less plausible analogues we substitute them in (20′) and obtain

$$L \frac{di}{dt} = E - Ki.$$

There is one small blemish. The frictional factor K is now associated with i instead of with v; speaking strictly, it becomes an electric

resistance factor. Therefore it is more appropriate and indeed customary in textbooks of electricity to call this factor r (for *resistance*, of course). With the substitution of r for K (20′) finally becomes

$$(20'') \qquad L\frac{di}{dt} = E - ri,$$

and our analogy is complete. Tidy minded we finish our tabulation.

	Analogy	
Electric resistance	$r \sim K$	Friction

Having found an analogy, or to moderate our claim, having found what we conjecture to be a sound analogy, we hasten to use it. The first step is to rewrite (20)–(22) on p. 190 in terms of v and K. As already noted, (20) becomes (20′). The initial condition becomes

$$\text{when} \quad t = 0, \quad x = 0, \quad v = 0.$$

And putting K/m for k in (22), it becomes

$$(22') \qquad v = \frac{mg}{K}(1 - e^{-(K/m)t}).$$

In short, the result of our investigation of free fall with friction may be expressed as follows:

If a phenomenon satisfies

$$(20') \qquad m\frac{dv}{dt} = mg - Kv$$

with the initial condition

$$t = 0, \quad x = 0, \quad v = 0,$$

then it satisfies

$$(22') \qquad v = \frac{mg}{K}(1 - e^{-(K/m)t}).$$

The road is now clear to speed to the consequence of our analogy. With the paired counterparts set before us thus:

$$\begin{array}{ccccc} m & dv/dt & mg & K & v \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow \\ L & di/dt & E & r & i \end{array}$$

we can immediately write:

If a phenomenon satisfies

$$L\frac{di}{dt} = E - ri \tag{20''}$$

with the initial condition

$$t = 0, \text{----}, i = 0$$

then it satisfies

$$i = \frac{E}{r}(1 - e^{-(r/L)t}). \tag{22''}$$

A trivial point: $x = 0$ has no counterpart; being extraneous it is cast aside without complaint. It is a lack of relevant, not a surplus of irrelevant, information which would be a cause for dissatisfaction.

It remains to ask the vital question: Is our analogy sound? It is. And what are our grounds for this assertion? As ever, experiment is the final arbiter. (22″) accords with the result of experiment; consequently, we accept (20″).

Earlier (using the less explicit notation) we showed that when $(k/m)t$ is large (22′) gives

$$v \approx \frac{mg}{K}$$

i.e., a falling body acquires a terminal or steady velocity. We must anticipate an analogous result for the flow of an electric current. When $(r/L)t$ is large (22″) gives

$$i \approx \frac{E}{r},$$

i.e., a flowing current acquires a terminal or steady intensity. Here is Ohm's Law, known to every schoolboy. We have additional grounds for accepting (20″) and for the soundness of our analogy.

Of course analogy is often misleading. Its importance is that it is often helpful. That we cannot give other examples of its role in finding new interpretations of old equations is lack of time, not material. Differential equations are powerful, for their interpretation is legion, and they speak with many tongues.

SECTION 5.4. WHAT IS A DIFFERENTIAL EQUATION?

5.4.1 Example

We have to solve the differential equation

$$\frac{dy}{dx} = -\frac{x}{y}. \tag{46}$$

I take the liberty to ask the reader to put down his pen, abstain from calculation, and THINK. What sort of geometric fact is expressed by equation (46)?

We are supposed to find a curve. Equation (46) says that the slope of the desired curve at the point (x, y) is $-x/y$. Now y/x is the slope of the line joining the origin to the point (x, y). Thus, very little knowledge of analytic geometry is sufficient to see what equation (46) really expresses: The desired curve is such that its tangent at any one of its points is perpendicular to the straight line joining that point to the origin, see Fig. 5.24. Do you know such a curve?

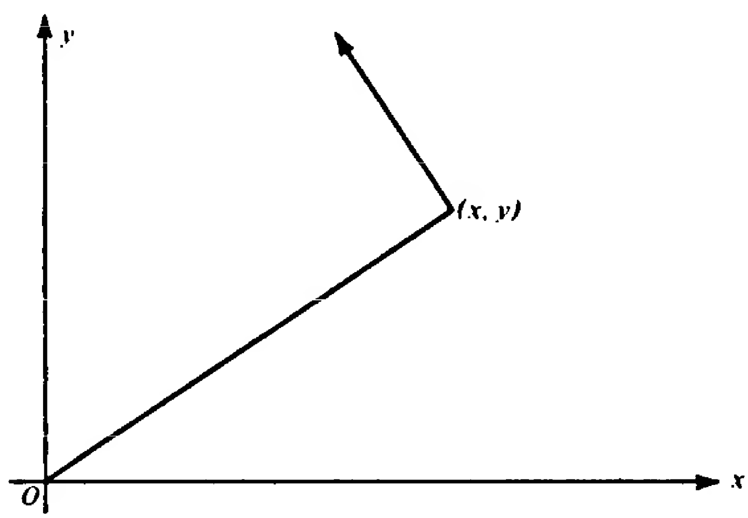

Figure 5.24 The tangent is perpendicular to the line joining the point of tangency to the origin.

To express the full import of the differential equation (46) we have to imagine, at each point (x, y) of the plane, the direction of the tangent

that (46) prescribes; this direction is shown by Fig. 5.25 at several points —it cannot be shown at all points. Do you know what curves have all along them the prescribed tangential directions?

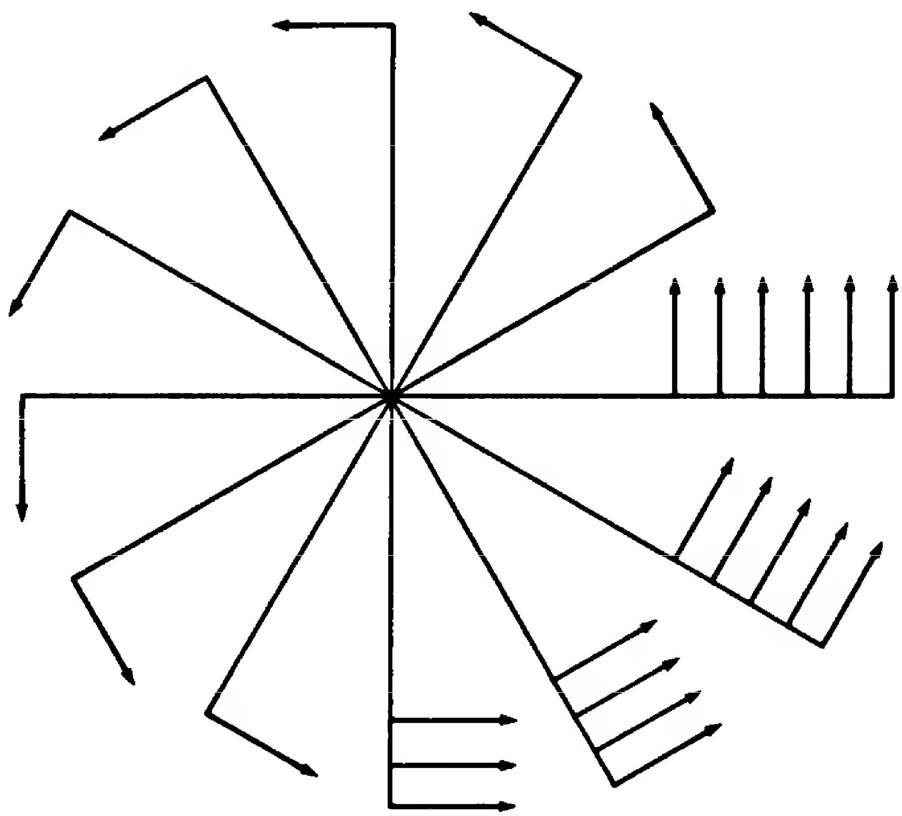

Figure 5.25 The rule prescribes the direction of the tangent at each point (with one exception).

The full import of equation (46) (the totality of directions it prescribes) is only hinted, but not completely represented, by Fig. 5.25. We can, however, obtain a complete representation by mechanical means: Let a rigid plate rotate about the origin; each of its points has a direction of motion and precisely the direction prescribed by equation (46). Have you any doubt what kind of curves satisfy equation (46)?

If there is any doubt in your mind let us calculate and solve the first order differential equation (46) by the method we have learned in this chapter. We can separate the variables:

$$y\,dy = -x\,dx,$$

integrate getting

$$\frac{y^2}{2} = -\frac{x^2}{2} + C,$$

and, introducing an appropriate notation for the constant of integration,

namely $C = r^2/2$ write the result in the form

$$x^2 + y^2 = r^2,$$

where r is an arbitrary positive constant. The differential equation (46) is satisfied by all circles whose common center is the origin, as we knew all along.

[Look critically at Figures 5.24 and 5.25. Do they express quite the same thing geometrically as the differential equation expresses in symbols? Is the caption of Fig. 5.25 quite correct?]

The next subsection 5.4.2 prepares us for a generalization of the example just considered.

5.4.2 Vector Fields

The steady flow of a river, which inspired many poets and the Greek philosopher Heraclitus, can also suggest important physical and mathematical ideas.

The flow is called steady if it does not vary in (i.e. is independent of) time. A steadily flowing river presents an unchanging view to the eye although it is in incessant motion. We would not detect any change in it even if we had some powerful instrument with which we could observe, without disturbing its flow, all details of its motion at once.

What are the essential details of the steady motion of a fluid? At every point the flow has a certain velocity, the velocity of the particle (a minute portion of the fluid) passing that point. The velocity of the flow does not vary in time, because the motion is steady, but it may vary from point to point. The velocity of the flow at any given point has magnitude and direction; it is a vector, it can be represented by a directed line segment. (Cf. Section 2.2.) To know the motion of the fluid completely we should know this vector, the velocity of the flow, at every point of it.

We arrived here at the important mathematico-physical concept of a vector field: a region of space to every point of which a certain vector, specified in magnitude and direction, is attached. The vector can be a force acting on the point, or the velocity of a speck of matter passing the point, and it can be geometrically represented as a directed line segment issuing from the point.

To do justice to the concept, just introduced, of a vector field, we should illustrate it with several examples drawn from different branches of physics. This would be easy as there are many such examples, but would take up more space than is available here. Yet we should not fail to look a little more at our example, the steady flow of the river.

The path of a particle in a steady flow is called a *streamline*. The velocity of the particle passing a certain point is the vector attached to

(issuing from) that point. Hence, the streamline, the path of the particle, is tangent to the vector at that point. The streamline is a curve which is tangent at each of its points to the vector of the field issuing from that point. See Fig. 5.26. See also Fig. 5.25; if we interpret it as a vector field, the streamlines are concentric circles.

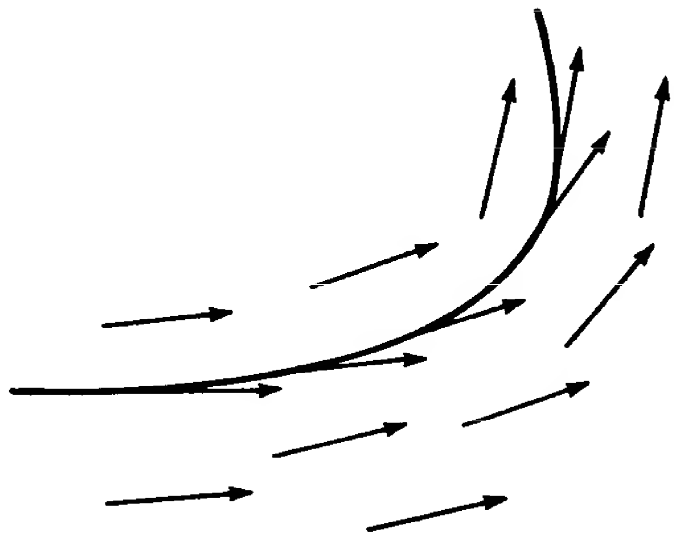

Figure 5.26 A streamline in a vector field

If the flow is steady, the field of velocity vectors and the system of streamlines remain unaffected by the progress of time. Looking at the vector field and its streamlines we do not notice any change. Yet, if we could distinguish the different particles of fluid from each other, we could observe incessant change: other and other particles pass the same point as time progresses.

We have here two aspects of a steady flow, one of unchanging persistence, the other of incessant change. To compare these aspects may be suggestive. It may suggest philosophical ideas to you if you are philosophically inclined.

Heraclitus was called the "Dark Philosopher"; his views of human affairs were sombre and his sayings obscure. Here are two enigmatic sentences attributed to him which are often quoted:

"You cannot look twice at the same river; for fresh waters are ever flowing in."

"We look and do not look at the same rivers; we are, and are not."

What is the intended meaning of these sentences? I do not venture to find that out. Yet I think that the originator of these sentences came pretty close to formulating the concept "steady flow of a fluid."

5.4.3 Direction Fields

In generalizing the example (46) discussed in subsection 5.4.1 we consider now the differential equation of first order

$$\frac{dy}{dx} = f(x, y), \tag{47}$$

where $f(x, y)$ stands for a given function.

The differential equation (47) prescribes the slope dy/dx at each point of the plane (or at each point of a certain region of the plane which we may call the "field".): The slope must be $f(x, y)$ at the point (x, y). We are required to find a curve (any curve, a particular curve, or every curve) that has the prescribed slope at each of its points.

We can imagine more vividly what the differential equation (47) tells us if we consider the plane in which x and y are rectangular coordinates as the free horizontal surface of a steadily flowing river.

What is given? A direction is given at each point of the field—take the direction of the fluid motion at that point (the direction of the attached vector, of the velocity of the flow) for the given direction.

What is the unknown? What are we required to find? A curve (any curve, or a particular curve, or every curve) that has the given direction at each of its points. Any fluid particle moving on the free horizontal surface of the river yields a solution of the differential equation: In fact, the path of the particle has at each of its points the given tangential direction, the direction of the flow at that point—and this path is a streamline.

Therefore, a differential equation of the first order of the form (47) can be conceived intuitively as a problem about the steady flow of a river: *Being given the direction of the flow at each point, find the streamlines.*

If the "general integral" is required, that is, the family of all solutions of the differential equation, we should find the system of all streamlines.

If less is required, just a "particular integral", that is, any solution of the differential equation, we should find some streamline.

If the solution of the differential equation should also satisfy an initial condition, we have to find the streamline that starts from a given initial point.

Yet we must add a comment. Elated by our new insight, we have expressed our physical interpretation of the differential equation (47) too briefly or, if you wish, incompletely. There is no objection to interpreting the curves satisfying the equation as streamlines in a steady flow. Then equation (47) does give us the slope $f(x, y)$ of the streamline passing through the point (x, y) at this point, but it does not give us quite the direction of the flow (of the velocity vector) at this point: It leaves open

the choice between the two possible directions in the line of given slope. (Here we may again quote Heraclitus: "The way up and the way down is one and the same.") Thus, in referring to the condition imposed by the differential equation on the streamlines, we should say specifically (and should have said above) "direction of an unoriented straight line" and not merely "direction." (Also the Figures 5.24 and 5.25 were slightly misleading in this respect.)

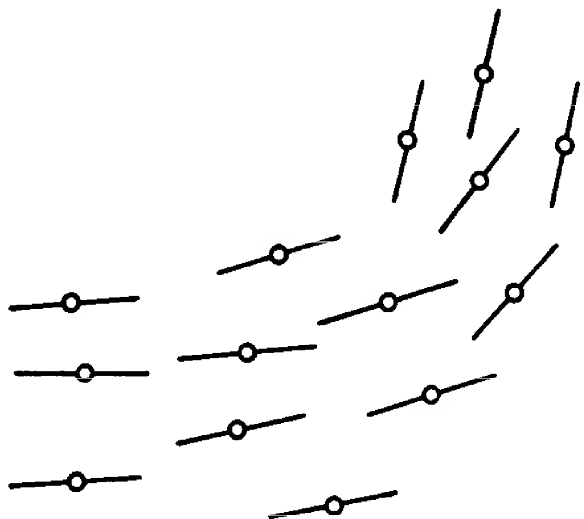

Figure 5.27 Direction field

In fact, the differential equation does not define a vector field: It does not tell us the magnitude of the vector attached to the point (x, y) and it does not even tell its direction completely since it leaves a choice between two alternatives. What the differential equation (47) really defines can be called, if the term "direction" is used in the above specified sense, a "direction field"; see Fig. 5.27 and compare it with Fig. 5.26.